

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 12:14:07 ; Search time 288.04 Seconds
(without alignments)
41.670 Million cell updates/sec

Title: US-09-508-147-10

Perfect score: 14
Sequence: 1 ctgctgacagcag 14

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	AA75956	Mouse p53CP 14 bp
2	14	100.0	14	AA75935	Mouse p53 DNA frag
3	14	100.0	19	AA75950	Mouse oligonucleot
4	14	100.0	20	AA75949	Mouse oligonucleot
5	14	100.0	20	AA75952	Mouse oligonucleot
6	14	100.0	20	AA75953	Mouse oligonucleot
7	14	100.0	20	AA75934	Mouse p53 DNA frag
8	13	92.9	20	AA75944	Mouse oligonucleot
9	13	92.9	20	AA75951	Mouse oligonucleot
10	13	92.9	21	AA75954	Mouse oligonucleot
11	13	92.9	131	AA75954	Human secreted pro

C 12	13	92.9	175	16	AA75934	Human gene signatu
C 13	13	92.9	318	20	AA75944	B. burgdorferi ant
C 14	13	92.9	366	21	AA75935	Human secreted pro
C 15	13	92.9	381	20	AA75949	B. burgdorferi ant
C 16	13	92.9	437	20	AA75950	Human secreted pro
C 17	13	92.9	468	22	AA75951	Human skeletal mus
C 18	13	92.9	528	22	AA75952	Probe #15889 for g
C 19	13	92.9	528	22	AA75953	Probe #22015 used
C 20	13	92.9	585	22	AA75934	Probe #6694 for ge
C 21	13	92.9	585	22	AA75944	Probe #8936 used t
C 22	13	92.9	599	21	AA75950	Human colon cancer
C 23	13	92.9	851	22	AA75951	Human CDNA clone (
C 24	13	92.9	900	21	AA75952	Human transmembran
C 25	13	92.9	903	21	AA75953	Arabidopsis thalia
C 26	13	92.9	1149	20	AA75935	Human transmembran
C 27	13	92.9	1276	21	AA75944	Human secreted pro
C 28	13	92.9	1305	20	AA75950	Human transmembran
C 29	13	92.9	1414	22	AA75951	Human colon cancer
C 30	13	92.9	1446	16	AA75952	E. coli trehalose
C 31	13	92.9	1446	16	AA75953	Trehalose phosphat
C 32	13	92.9	1446	17	AA75934	E. coli trehalose
C 33	13	92.9	1450	18	AA75935	E. coli trehalose
C 34	13	92.9	1482	13	AA75944	ACC synthetase gen
C 35	13	92.9	1597	21	AA75950	Arabidopsis thalia
C 36	13	92.9	1600	21	AA75951	Arabidopsis thalia
C 37	13	92.9	1683	22	AA75952	Human CDNA encodin
C 38	13	92.9	1697	21	AA75935	Human transmembran
C 39	13	92.9	1703	12	AA75944	Clone PAC1 encodi
C 40	13	92.9	1703	19	AA75950	Zucchini ACC synth
C 41	13	92.9	1703	22	AA75951	Zucchini 1-aminocy
C 42	13	92.9	1703	22	AA75952	Zucchini ACC synth
C 43	13	92.9	1711	21	AA75953	Human DNA encodin
C 44	13	92.9	1734	20	AA75944	Human secreted pro
C 45	13	92.9	1755	22	AA75954	Human Sec9 DNA seq

ALIGNMENTS

RESULT 1	
AA75956	AA75956 standard; DNA; 14 BP.
ID	AA75956
XX	
AC	AA75956;
XX	
DT	29-JUL-1999 (first entry)
XX	
DE	Mouse p53CP 14 bp motif.
XX	
KW	Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KW	DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KW	tumour cell growth inhibition; genome guardian; differentiation;
KW	senescence; angiogenesis; ss.
OS	Mus musculus.
XX	
PN	W09925820-A1.
PD	27-MAY-1999.
PF	10-NOV-1998; 98WO-US23992.
XX	
PR	17-NOV-1997; 97US-0065740.
PA	(WARN) WARNER LAMBERT CO.
XX	
PI	Bian J, Sun Y;
DR	WPI, 1999-347468/29.
PT	New p53CP protein that specifically binds to the p53 consensus
XX	binding sites, useful for treating p53 associated disorders

PS Disclosure; Page 25; 37pp; English.
 XX
 CC The present invention describes a p53 competing protein designated p53CP
 CC (40 kD) that specifically binds to the p53 consensus binding sites.
 CC The p53CP protein is useful for p53 inactivation and regulation during
 CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
 CC growth inhibition, genome guardian, differentiation, senescence, and
 CC angiogenesis.
 XX
 SQ Sequence 14 BP; 3 A; 3 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgtgttgacacag 14
 |||||
 DB 1 ctgtgttgacacag 14

RESULT 2
 AAX75935
 ID AAX75935 standard; DNA; 14 BP.
 XX
 AC AAX75935;
 XX
 DT 29-JUL-1999 (first entry)
 XX
 DE Mouse p53 DNA fragment SEQ ID NO:10.
 XX
 KW Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
 KW DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
 KW tumour cell growth inhibition; genome guardian; differentiation;
 KW senescence; angiogenesis; ss.
 XX
 OS Mus musculus.
 XX
 PN WO9925820-A1.
 XX
 PD 27-MAY-1999.
 XX
 PF 10-NOV-1998; 98WO-US23992.
 XX
 PR 17-NOV-1997; 97US-0065740.
 XX
 PA (WARN) WARNER LAMBERT CO.
 XX
 PI Brian J, Sun Y;
 XX
 DR WPI: 1999-347468/29.
 XX
 PT New p53CP protein that specifically binds to the p53 consensus
 PT binding sites, useful for treating p53 associated disorders
 XX
 PS Claim 3; Page 28; 37pp; English.
 XX
 CC The present invention describes a p53 competing protein designated p53CP
 CC (40 kD) that specifically binds to the p53 consensus binding sites.
 CC The p53CP protein is useful for p53 inactivation and regulation during
 CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
 CC growth inhibition, genome guardian, differentiation, senescence, and
 CC angiogenesis. The present sequence represents a mouse p53 DNA fragment
 CC from the present invention.
 XX
 SQ Sequence 14 BP; 3 A; 3 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgtgttgacacag 14

DB 1 ctgtgttgacacag 14
 |||||
 DB 1 ctgtgttgacacag 14

RESULT 3
 AAX75950
 ID AAX75950 standard; DNA; 19 BP.
 XX
 AC AAX75950;
 XX
 DT 29-JUL-1999 (first entry)
 XX
 DE Mouse oligonucleotide T3SF2.
 XX
 KW Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
 KW DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
 KW tumour cell growth inhibition; genome guardian; differentiation;
 KW senescence; angiogenesis; ss.
 XX
 OS Mus musculus.
 XX
 PN WO9925820-A1.
 XX
 PD 27-MAY-1999.
 XX
 PF 10-NOV-1998; 98WO-US23992.
 XX
 PR 17-NOV-1997; 97US-0065740.
 XX
 PA (WARN) WARNER LAMBERT CO.
 XX
 PI Brian J, Sun Y;
 XX
 DR WPI: 1999-347468/29.
 XX
 PT New p53CP protein that specifically binds to the p53 consensus
 PT binding sites, useful for treating p53 associated disorders
 XX
 PS Disclosure; Page 22; 37pp; English.
 XX
 CC The present invention describes a p53 competing protein designated p53CP
 CC (40 kD) that specifically binds to the p53 consensus binding sites.
 CC The p53CP protein is useful for p53 inactivation and regulation during
 CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
 CC growth inhibition, genome guardian, differentiation, senescence, and
 CC angiogenesis.
 XX
 SQ Sequence 19 BP; 3 A; 4 C; 4 G; 8 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 19;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgtgttgacacag 14
 |||||
 DB 2 ctgtgttgacacag 15

RESULT 4
 AAX75949
 ID AAX75949 standard; DNA; 20 BP.
 XX
 AC AAX75949;
 XX
 DT 29-JUL-1999 (first entry)
 XX
 DE Mouse oligonucleotide T3SF1.
 XX
 KW Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
 KW DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
 KW tumour cell growth inhibition; genome guardian; differentiation;
 KW senescence; angiogenesis; ss.

XX Mus musculus.
OS
XX WO9925820-A1.
PN
XX 27-MAY-1999.
PD
XX 10-NOV-1998; 98WO-US23992.
PF
XX 17-NOV-1997; 97US-0065740.
PR
XX (WARN) WARNER LAMBERT CO.
PA
XX
PI Bian J, Sun Y;
XX
DR WPI: 1999-347468/29.
XX
XX New p53CP protein that specifically binds to the p53 consensus
PT binding sites, useful for treating p53 associated disorders
XX
XX
PS Disclosure: Page 22; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kd) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
XX
SQ Sequence 20 BP; 3 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ctgcttgacacag 14
|||||
-DB 3 ctgcttgacacag 16

RESULT 5
AAK75952
ID AAK75952 standard; DNA: 20 BP.
XX
XX AAK75952;
AC
XX
DT 29-JUL-1999 (first entry)
XX
DE Mouse oligonucleotide T3SF4.
XX
KM Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KM DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KM tumour cell growth inhibition; genome guardian; differentiation;
KM senescence; angiogenesis; ss.
XX
OS Mus musculus.
XX
PN WO9925820-A1.
XX
PD 27-MAY-1999.
XX
PF 10-NOV-1998; 98WO-US23992.
XX
PR 17-NOV-1997; 97US-0065740.
XX
PA (WARN) WARNER LAMBERT CO.
XX
PI Bian J, Sun Y;
XX
DR WPI: 1999-347468/29.
XX
PT New p53CP protein that specifically binds to the p53 consensus

PT binding sites, useful for treating p53 associated disorders
XX
XX Disclosure: Page 22; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kd) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
XX
SQ Sequence 20 BP; 3 A; 3 C; 5 G; 9 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ctgcttgacacag 14
|||||
DB 3 ctgcttgacacag 16

RESULT 6
AAK75953
ID AAK75953 standard; DNA: 20 BP.
XX
XX AAK75953;
AC
XX
DT 29-JUL-1999 (first entry)
XX
DE Mouse oligonucleotide T3SF5.
XX
KM Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KM DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KM tumour cell growth inhibition; genome guardian; differentiation;
KM senescence; angiogenesis; ss.
XX
OS Mus musculus.
XX
PN WO9925820-A1.
XX
PD 27-MAY-1999.
XX
PF 10-NOV-1998; 98WO-US23992.
XX
PR 17-NOV-1997; 97US-0065740.
XX
PA (WARN) WARNER LAMBERT CO.
XX
PI Bian J, Sun Y;
XX
DR WPI: 1999-347468/29.
XX
PT New p53CP protein that specifically binds to the p53 consensus
PT binding sites, useful for treating p53 associated disorders
XX
XX Disclosure: Page 22; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kd) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
XX
SQ Sequence 20 BP; 3 A; 3 C; 4 G; 10 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcttgacacag 14
|||||
Db 4 ctgcttgacacag 17

RESULT 7

AAK75934 standard; DNA: 20 BP.

AAK75934;

29-JUL-1999 (first entry)

Mouse p53 DNA fragment SEQ ID NO:8.

Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
tumour cell growth inhibition; genome guardian; differentiation;
senescence; angiogenesis; ss.

Mus musculus.

WO9925820-A1.

27-MAY-1999.

10-NOV-1998; 98WO-US23992.

17-NOV-1997; 97US-0065740.

(WARN) WARNER LAMBERT CO.

Bian J, Sun Y;

WPI: 1999-347468/29.

New p53CP protein that specifically binds to the p53 consensus
binding sites, useful for treating p53 associated disorders

Claim 2; Page 28; 37pp; English.

The present invention describes a p53 competing protein designated p53CP
(40 kD) that specifically binds to the p53 consensus binding sites.
The p53CP protein is useful for p53 inactivation and regulation during
human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
growth inhibition, genome guardian, differentiation, senescence, and
angiogenesis. The present sequence represents a mouse p53 DNA fragment
from the present invention.

Sequence 20 BP; 3 A; 4 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 14; DB 20; Length 20;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcttgacacag 14
|||||
Db 4 ctgcttgacacag 17

RESULT 8

AAK75944/C standard; DNA: 20 BP.

AAK75944;

29-JUL-1999 (first entry)

Mouse oligonucleotide T3SF SEQ ID NO:9.

Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
DNA binding site; p53 competing protein; carcinogenesis; apoptosis;

KW tumour cell growth inhibition; genome guardian; differentiation;
senescence; angiogenesis; ss.

Mus musculus.

WO9925820-A1.

27-MAY-1999.

10-NOV-1998; 98WO-US23992.

17-NOV-1997; 97US-0065740.

(WARN) WARNER LAMBERT CO.

Bian J, Sun Y;

WPI: 1999-347468/29.

New p53CP protein that specifically binds to the p53 consensus
binding sites, useful for treating p53 associated disorders

Disclosure; Page 7; 37pp; English.

The present invention describes a p53 competing protein designated p53CP
(40 kD) that specifically binds to the p53 consensus binding sites.
The p53CP protein is useful for p53 inactivation and regulation during
human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
growth inhibition, genome guardian, differentiation, senescence, and
angiogenesis.

Sequence 20 BP; 5 A; 8 C; 4 G; 3 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgcttgacacag 14
|||||
Db 20 TTGCTTGACACAG 8

RESULT 9

AAK75951 standard; DNA: 20 BP.

AAK75951;

29-JUL-1999 (first entry)

Mouse oligonucleotide T3SF3.

Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
tumour cell growth inhibition; genome guardian; differentiation;
senescence; angiogenesis; ss.

Mus musculus.

WO9925820-A1.

27-MAY-1999.

10-NOV-1998; 98WO-US23992.

17-NOV-1997; 97US-0065740.

(WARN) WARNER LAMBERT CO.

Bian J, Sun Y;

WPI: 1999-347468/29.


```

AC AAT26534;
XX 10-OCT-1996 (first entry)
XX Human gene signature HUMGS08778.
XX
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX human; cloning; mapping; non-biased library; diagnosis; detection;
XX cell typing; abnormal cell function; ss.
XX Homo sapiens.
XX W09514772-A1.
XX
XX 01-JUN-1995.
XX
XX 11-NOV-1994; 94MO-JP01916.
XX
XX 12-NOV-1993; 93JP-0355504.
XX
XX (MATS/) MATSUBARA K.
XX (OKUBO/) OKUBO K.
XX
XX Matsubara K, Okubo K;
XX
XX WPI: 1995-206931/27.
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX reflects relative abundance of corresp. mRNA in specific human
XX tissues
XX
XX Claim 1; Page 2110; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX given in AAT19001-T26837 and which is able to hybridize to part of
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS
XX sequences) as a means of diagnosing abnormal cell function or for
XX recognising different cell types.
XX
XX Sequence 175 BP; 67 A; 24 C; 32 G; 50 T; 2 other;
XX
XX
XX Query Match 92.9%; Score 13; DB 16; Length 175;
XX Best Local Similarity 92.9%; Pred. No. 2.2e+02;
XX Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
Oy 1 ctgtgctgaacag 14
Db 38 CTGCTTGACAG 25

```

```

XX
XX Borrelia burgdorferi.
XX
XX W09859071-A1.
XX
XX 30-DEC-1998.
XX
XX 18-JUN-1998; 98MO-US12718.
XX
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050359.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMUNE INC.
XX
XX Chol GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX WPI: 1999-189980/16.
XX P-PSDB: AAY19997.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of diseases
XX caused by Borrelia, particularly Lyme disease
XX
XX Claim 1; Page 155; 275pp; English.
XX
XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus.
XX
XX Sequence 318 BP; 126 A; 49 C; 51 G; 92 T; 0 other;
XX
XX
XX Query Match 92.9%; Score 13; DB 20; Length 318;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Oy 1 ctgtgctgaacag 13
Db 173 CTGCTTGACAG 161

```

```

XX
XX AATC01913 standard; cDNA; 366 BP.
XX
XX AATC01913;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 1911.
XX
XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX

```

PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR MPI: 2000-500381/45.
 DR P-PSDB; AAC01907.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 1911; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 CC
 SQ Sequence 366 BP; 85 A; 77 C; 122 G; 82 T; 0 other;

Query Match 92.9%; Score 13; DB 21; Length 366;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgcttgacag 14
 ||||||||||||
 DB 236 TTGCTTGACAG 224

RESULT 15

AAK61693/c
 ID AAK61693 standard; DNA; 381 BP.

XX
 AC AAK61693;

DT 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein coding sequence, f805.nt.

XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

OS Borrelia burgdorferi.

XX
 PN W09859071-A1.

XX
 PD 30-DEC-1998.

PF 18-JUN-1998; 98WO-US12718.

XX
 PR 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

PR 22-JUL-1997; 97US-0053377.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI (MEDI-) MEDIMUNE INC.

XX
 PI Chol GH, Erwin AL, Hanson MS, Lathigra R;

DR MPI: 1999-189980/16.

DR P-PSDB; AAT19996.

PR caused by Borrelia, particularly Lyme disease
 XX
 PS Claim 1; Page 155; 275bp; English.
 XX
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.
 CC
 SQ Sequence 381 BP; 150 A; 59 C; 62 G; 110 T; 0 other;

Query Match 92.9%; Score 13; DB 20; Length 381;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttgacag 13
 ||||||||||||
 DB 236 CTTGCTTGACAG 224

Search completed: December 8, 2001, 12:14:08
 Job time: 9648 sec

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 09:28:40 ; Search time 2889.21 Seconds
(without alignments)
74.386 Million cell updates/sec

Title: US-09-508-147-8

Perfect score: 20

Sequence: 1 ggcttgcttgatgaacaggtc 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inu:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.4	87.0	501	10	AM032968 EST276527
2	17	85.0	641	11	BF327660 RC2-BN012
3	16.8	84.0	273	10	AA684691 EST105126
4	16.8	84.0	298	10	AA686109 EST109246
5	16.8	84.0	330	10	BB312230 BB312230
6	16.8	84.0	398	11	BG663609 BSA10246
7	16.8	84.0	426	11	H35586 EST10188 R
8	16.8	84.0	452	10	BE725984
9	16.8	84.0	526	10	AV393184
10	16.8	84.0	533	10	AV623174
11	16.8	84.0	640	13	A2273960
12	16.8	84.0	647	13	AQ994742

C 13	16.8	84.0	652	11	BG855209
C 14	16.8	84.0	675	10	BE381989
C 15	16.8	84.0	684	10	AA914280
C 16	16.8	84.0	927	11	BG255332
C 17	16.8	84.0	1013	11	B1146463
C 18	16.4	82.0	187	10	AA866904
C 19	16.4	82.0	253	10	BB392933
C 20	16.4	82.0	264	13	AZ003937
C 21	16.4	82.0	335	10	AA281460
C 22	16.4	82.0	467	11	W96351
C 23	16.4	82.0	503	10	AW030956
C 24	16.4	82.0	506	11	BE515523
C 25	16.4	82.0	532	11	BF176526
C 26	16.4	82.0	533	10	AA625367
C 27	16.4	82.0	596	13	AZ902400
C 28	16	80.0	231	10	AA265067
C 29	16	80.0	306	11	BF456636
C 30	16	80.0	438	10	A1154122
C 31	16	80.0	485	10	AA670946
C 32	16	80.0	486	13	AA0873059
C 33	16	80.0	501	11	BF723243
C 34	16	80.0	527	10	BE689712
C 35	16	80.0	604	13	AQ503365
C 36	16	80.0	735	13	AQ162543
C 37	16	80.0	769	11	B1221887
C 38	15.8	79.0	122	10	A1007600
C 39	15.8	79.0	274	10	AA684672
C 40	15.8	79.0	293	10	AA685469
C 41	15.8	79.0	316	11	BF811376
C 42	15.8	79.0	319	10	AA686473
C 43	15.8	79.0	407	10	A1821281
C 44	15.8	79.0	437	13	AQ517507
C 45	15.8	79.0	440	11	T81254

ALIGNMENTS

RESULT 1
AM032968/c 501 bp mRNA EST 18-MAY-2001
LOCUS EST276527 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION C1EC117N7, mRNA sequence.
AM032968
ACCESSION AM032968.1 GI:5891724
VERSION EST.
KEYWORDS tomato.
SOURCE Lycopersicon esculentum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 501)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Jiang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Rönning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J. Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..501
/organism="Lycopersicon esculentum"
/cultivar="TMA96"
/db_xref="taxon:4081"
/clone="C1EC117N7"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"

/lab.host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato callus EST library"

BASE COUNT 167 a 102 c 79 g 153 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 501;
Best Local Similarity 94.7%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ggcctgttgacaggtc 20
|||||
Db 163 GGCCTTGACAGGATC 145

RESULT 2
LOCUS BF327660 641 bp mRNA EST 22-NOV-2000
DEFINITION RC3-BN0123-170300-011-h11 BN0123 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF327660
VERSION BF327660.1 GI:11298408
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 641)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bata, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=RC2&t2=RC2-BN0123-170300-011-h11&t3=2000-03-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 32.
Location/Qualifiers

FEATURES
SOURCE 1..641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0123"
/dev_stage="Adult"

/note="Organ: breast, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 189 a 117 c 120 g 215 t
ORIGIN

Query Match 85.0%; Score 17; DB 11; Length 641;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 ctgtcttgacaggtc 20
|||||
Db 577 CTTGCTTGACAGGGTC 593

RESULT 3
LOCUS AA684691/c 273 bp mRNA EST 13-JAN-1998
DEFINITION EST105126 Rat PC-12 cells, untreated Rattus sp. cDNA 5' end similar to Secretogranin I, mRNA sequence.
ACCESSION AA684691
VERSION AA684691.1 GI:2671289
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 273)
AUTHORS Lee, N. H., Weinstock, K. G., Kirkness, E. F., Earle-Hughes, J. A., Fuldner, R. A., Marmaras, S., Glodex, A., Gocayne, J. D., Adams, M. D., Kerlavage, A. R., Fraser, C. M. and Venter, J. C.
TITLE Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment
Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
COMMENT Other ESTs: TC45246
95396786

JOURNAL The Institute for Genomic Research
MEDLINE 9712, Medical Center Drive, Rockville, MD 20850, USA
COMMENT Contact: Lee, NH
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igr.org
Seq primer: M13 Reverse.

FEATURES
SOURCE Location/Qualifiers
1..273
/organism="Rattus sp."
/db_xref="ATCC (Inhost):2000329"
/db_xref="taxon:10118"
/clone_lib="Rat PC-12 cells, untreated"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; poly(A)+ RNA was purified from untreated PC12 cells cultured for 9 days. cDNA was constructed using an oligo-dT primer and directionally cloned using the lambda ZAP II Vector Kit by Stratagene"

BASE COUNT 96 a 54 c 87 g 36 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 273;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ggcctgttgacaggtc 20
|||||
Db 101 GGCCTTGACAGGGTC 82

RESULT 4
LOCUS AA686109/c 296 bp mRNA EST 03-APR-1998
DEFINITION EST109246 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA clone RPNAP27 5' end similar to Secretogranin I, mRNA sequence.
ACCESSION AA686109
VERSION AA686109.1 GI:2672707
KEYWORDS EST.
SOURCE Rattus sp.

ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 298)
Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner
,R.A., Marmaras,S., Glodok,A., Gocayne,J.D., Adams,M.D., Kerlavage
,A.R., Fraser,C.M. and Venter,J.C.
Comparative expressed-sequence-tag analysis of differential gene
expression profiles in PC-12 cells before and after nerve growth
factor treatment
Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
JOURNAL MEDLINE 95396786
COMMENT Other ESTs: TC45246
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13 Reverse
FEATURES Location/Qualifiers
source 1..298
/organism="Rattus sp."
/db_xref="ATCC (Inhost):2003682"
/db_xref="taxon:10118"
/clone="RPNP27"
/clone_1lb="Rat PC-12 cells, NGF-treated (9 days)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; poly(A)+ RNA was purified from 9-day NGF treated
PC12 cells. cDNA was constructed using an oligo-dT primer
and directionally cloned using the lambda Zap II Vector
kit by Stratagene"

BASE COUNT 103 a 67 c 85 g 42 t 1 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 298;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggagctgctgacagcgc 20
||||| 1 |||||||||
Db 45 GGGCTTTCATGACAGGCTC 26

RESULT 5
BB312230 330 bp mRNA EST 10-JUL-2000
LOCUS BB312230 RIKEN full-length enriched, adult male corpora
DEFINITION quadrigemina Mus musculus cDNA clone B23033B10 3', mRNA sequence.
ACCESSION BB312230
VERSION BB312230.1 GI:9012935
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 330)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arahara,T., Carininci
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hitozane,T., Hori,F., Iehi,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,U., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Oto,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomioka,N., Toyu
,T., Tsunoda,Y., Watanabe,A., Watanabe,S., Yamamura,T., Yamada,I.,
,Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)

COMMENT Contact: Yoshitake Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: http://genome.resc.riken.go.jp/
Carininci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermocatalysis of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carininci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carininci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.ritc.riken.go.jp>) for
further details.
FEATURES Location/Qualifiers
source 1..330
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B23033B10"
/clone_1lb="RIKEN full-length enriched, adult male corpora
quadrigemina"
/sex="male"
/tissue="type="corpora quadrigemina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATTCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
transcriptase and subsequently thermo-activated reverse
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATTAATTCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 73 a 78 c 85 g 94 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 330;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggagctgctgacagcgc 20
||||| 1 |||||||||
Db 35 GGGCTTTCATGACAGGCTC 54

RESULT 6
BG663609 398 bp mRNA EST 30-APR-2001
LOCUS BG663609 RIKEN full-length enriched, adult male corpora
DEFINITION DRABALB12 Rat DRG Library Rattus norvegicus cDNA clone DRABALB12
5', mRNA sequence.
ACCESSION BG663609
VERSION BG663609.1 GI:13885531
KEYWORDS EST.

SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 398)
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G., Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and Zhang X.
Distinct gene expression profiles of rat dorsal root ganglion induced by peripheral nerve axotomy
TITLE
JOURNAL Unpublished (2001)
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzeg@ion.ac.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source
1.398
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRRAALB12"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"

BASE COUNT 139 a 87 c 113 g 59 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 398;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggagctgctgaacaggc 20
||||| | ||||| |||||
Db 79 GGGCTTTCATGAACAGGCTC 60

RESULT 7
LOCUS H35586 426 bp mRNA EST 02-APR-1998
DEFINITION EST110188 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA
clone RPNC39 similar to Secretogranin I, mRNA sequence.
ACCESSION H35586
VERSION H35586.1 GI:981003
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus

REFERENCE 1 (bases 1 to 426)
Lee,N.H., Weinstein,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner ,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D., Kerlavage ,A.R., Fraser,C.M. and Venter,J.C.
Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
MEDLINE 95396786

COMMENT Other_ESTs: TC72
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability please contact the TIGR Database (tdbinfo@tdb.tigr.org) TC (Tentative Consensus) numbers represent assemblies of ESTs.

FEATURES
source
1.426
Location/Qualifiers
/organism="Rattus sp."
/db_xref="ATCC (inhost):2004374"
/db_xref="taxon:10118"
/clone="RPNC39"
/clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
/note="Vector: pBluescript SK-; Site1: EcoRI; Site2: XhoI; poly(A)+ RNA was purified from 9-day NGF treated PC12 cells. cDNA was constructed using an oligo-dT primer and directionally cloned using the Lambda Zap II Vector kit by Stratagene"

BASE COUNT 142 a 83 c 120 g 74 t 7 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 426;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggagctgctgaacaggc 20
||||| | ||||| |||||
Db 46 GGGCTTTCATGAACAGGCTC 27

RESULT 8
LOCUS BE725984 452 bp mRNA EST 14-SEP-2000
DEFINITION 894087607.Y1 C. reinhardtii CC-1690, normalized, Lambda zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BE725984
VERSION BE725984.1 GI:10127280
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 452)
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Sillflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants, project phase 2
Unpublished (2000)
JOURNAL Contact: Charles Hauser
CDMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1.452
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 219r"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda zap II"
/note="Vector: pBluescript II SK-; Site1: EcoRI; Site2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in

ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 81 a 160 c 154 g 57 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 452;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gggctgtctgaacaggc 20
|||||
DB 233 GGGCCTGCTTGACAGGGC 214

RESULT 9
AV393184/C 526 bp mRNA EST 29-SEP-2000
LOCUS AV393184 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
DEFINITION CDNA clone CM097h07_r, mRNA sequence.
ACCESSION AV393184
VERSION AV393184.1 GI:6547400
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 526)
Asanizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
A large scale structural analysis of cDNAs in a unicellular green alga, Chlamydomonas reinhardtii. I. Generation of 3433 non-redundant expressed sequence tags
DNA Res. 6 (6), 369-373 (1999)

JOURNAL MEDLINE 20152986
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source 1..526
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone_1ib="CM097h07_r"
/clone_1ib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 88 a 196 c 167 g 75 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 526;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gggctgtctgaacaggc 20
|||||
DB 466 GGGCCTGCTTGACAGGGC 447

RESULT 10
AV623174/C 533 bp mRNA EST 15-DEC-2000
LOCUS AV623174 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LC059h12_r 5', mRNA sequence.

ACCESSION AV623174
VERSION AV623174.1 GI:10772351
KEYWORDS EST
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 533)
Asanizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.
Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)

JOURNAL MEDLINE 20539644
Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source 1..533
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone_1ib="LC059h12_r"
/clone_1ib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

BASE COUNT 106 a 171 c 180 g 76 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 533;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gggctgtctgaacaggc 20
|||||
DB 42 GGGCCTGCTTGACAGGGC 23

RESULT 11
AZ273960/C 640 bp DNA GSS 26-JUL-2000
LOCUS RPCI-23-167B5_TV RPCI-23 Mus musculus genomic clone RPCI-23-167B5,
DEFINITION DNA sequence.
ACCESSION AZ273960.1 GI:9487577
VERSION AZ273960.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 640)
Zhao,S., Niernan,W., Feldbiyum,T., Malek,J., Shatsman,S., Akintet ,B., Levins,M., McGinn,S., Tsegaye,G., Geer,K., Kiol,M., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-167B5_TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

JOURNAL MEDLINE
COMMENT
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html)
Plate: 167 row: B column: 5
Seq primer: T7
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1..640
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-167B5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 188 a 133 c 139 g 180 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 640;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 1 gggtgtgtgacagaggtc 20
||||| ||| |||||
Db 231 GGGCTTGTGGACAGGCTC 212

RESULT 12
LOCUS A0994742 647 bp DNA GSS 24-FEB-2000
DEFINITION RPCI-23-383H6.TV RPCI-23 Mus musculus genomic clone RPCI-23-383H6,
DNA sequence.
ACCESSION A0994742
VERSION A0994742.1 GI:7069839
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 647)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akintet,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-383H6.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html)
Plate: 383 row: H column: 6
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..647
/organism="Mus musculus"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-383H6"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 186 a 133 c 143 g 183 t 2 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 647;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 1 gggtgtgtgacagaggtc 20
||||| ||| |||||
Db 221 GGGCTTGTGGACAGGCTC 202

RESULT 13
LOCUS BG855209 652 bp mRNA EST 29-MAY-2001
DEFINITION 1024042B01.y2 C. reinhardtii CC-1690, normalized, lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG855209
VERSION BG855209.1 GI:14236393
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadales.
1 (bases 1 to 652)
McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; Project Phase 2
Unpublished (2000)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1..652
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, lambda Zap
II"

FEATURES
source
Location/Qualifiers
1..647
/organism="Mus musculus"

BASE COUNT 118 a 222 c 213 g 97 t 2 others

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 09:30:10 ; Search time 136.23 Seconds
(without alignments)
33.249 Million cell updates/sec

Title: US-09-508-147-8

Perfect score: 20
Sequence: 1 ggggttgcttgacacggggtc 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCYUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.2	76.0	2870	1	US-08-468-036-28 Sequence 28, Appl
2	15.2	76.0	2870	2	US-08-376-843-28 Sequence 28, Appl
3	14.8	74.0	712	1	US-08-276-452A-72 Sequence 72, Appl
4	14.8	74.0	712	2	US-08-798-744-72 Sequence 72, Appl
5	14.8	74.0	3735	4	US-08-975-762-43 Sequence 43, Appl
6	14.8	74.0	3735	4	US-09-295-028-43 Sequence 43, Appl
7	14.8	74.0	3735	4	US-09-106-582-43 Sequence 43, Appl
8	14.4	72.0	2220	2	US-08-932-376A-3 Sequence 3, Appl
9	14.4	72.0	3688	6	5248670-4 Patient No. 5248670
10	14.4	72.0	246240	2	US-08-724-394A-20 Sequence 20, Appl
11	14.4	72.0	246240	2	US-08-724-394A-21 Sequence 21, Appl
12	14.4	72.0	246240	2	US-08-724-394A-22 Sequence 22, Appl
13	14.2	71.0	79	3	US-09-039-555B-5 Sequence 5, Appl
14	14.2	71.0	343	1	US-08-473-020A-6 Sequence 6, Appl
15	14.2	71.0	2638	1	US-08-306-691B-46 Sequence 46, Appl
16	14.2	71.0	4245	4	US-09-276-531-16 Sequence 16, Appl
17	14.2	71.0	24417	2	US-08-846-762-1 Sequence 16, Appl
18	13.8	69.0	657	4	US-08-998-416-1061 Sequence 1061, Ap
19	13.8	69.0	1869	4	US-09-350-268-1 Sequence 1, Appl
20	13.8	69.0	2088	3	US-09-032-365A-64 Sequence 64, Appl
21	13.8	69.0	3046	1	US-08-525-507-7 Sequence 7, Appl
22	13.8	69.0	3919	2	US-08-866-650-4 Sequence 4, Appl
23	13.8	69.0	3919	2	US-09-021-287-4 Sequence 4, Appl
24	13.8	69.0	3919	2	US-09-240-473-4 Sequence 4, Appl
25	13.8	69.0	4771	2	US-08-866-650-2 Sequence 2, Appl
26	13.8	69.0	4771	2	US-09-021-287-2 Sequence 2, Appl
27	13.8	69.0	4771	4	US-09-240-473-2 Sequence 2, Appl

28	13.8	69.0	4935	2	US-08-631-097-3 Sequence 3, Appl
c 29	13.8	69.0	4985	4	US-08-998-416-7 Sequence 7, Appl
c 30	13.8	69.0	5145	3	US-08-991-406-1 Sequence 1, Appl
c 31	13.8	69.0	5285	2	US-08-609-042A-29 Sequence 29, Appl
c 32	13.8	69.0	5285	4	US-09-170-996-29 Sequence 29, Appl
c 33	13.8	69.0	5866	4	US-08-810-712-9 Sequence 9, Appl
34	13.8	69.0	6416	4	US-09-136-574A-2 Sequence 2, Appl
35	13.8	69.0	6693	2	US-08-147-777-2 Sequence 2, Appl
36	13.8	69.0	6693	3	US-08-452-872-2 Sequence 2, Appl
37	13.8	69.0	6693	5	PCT-US93-03985-2 Sequence 2, Appl
38	13.8	69.0	7498	3	US-08-816-693A-1 Sequence 1, Appl
39	13.8	69.0	7498	4	US-08-885-291-1 Sequence 1, Appl
40	13.8	69.0	7498	4	US-09-496-672-1 Sequence 1, Appl
c 41	13.8	69.0	7573	1	US-08-287-959-2 Sequence 2, Appl
c 42	13.8	69.0	10747	2	US-08-147-777-1 Sequence 1, Appl
c 43	13.8	69.0	10747	3	US-08-452-872-1 Sequence 1, Appl
44	13.8	69.0	10747	5	PCT-US93-03985-1 Sequence 1, Appl
45	13.8	69.0	24979	2	US-08-147-777-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-468-036-28
; Sequence 28, Application US/08468036
; Patent No. 5728806
; GENERAL INFORMATION:
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins that
; INTERACT WITH CASEIN KINASE I
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.25, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468, 036
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5728806and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6500
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2870 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-036-28
Query Match 76.0%; Score 15.2; DB 1; Length 2870;
Best Local Similarity 85.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 999ctgtctgacagggtc 20
|||
DB 2372 GGCCTGCTGAACGGGATC 2391

RESULT 2

US-08-376-843-28
Sequence 28, Application US/08376843
Patent No. 5846764
GENERAL INFORMATION:
APPLICANT: Demaggio, Anthony J.
APPLICANT: Hoekstra, Weri F.
TITLE OF INVENTION: Materials and Methods Relating to Proteins
TITLE OF INVENTION: that interact with Casein Kinase I
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/376.843
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,605
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: No. 5846764 and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31784
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 2870 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-376-843-28

Query Match 76.0%; Score 15.2; DB 2; Length 2870;
Best Local Similarity 85.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 999ctgtctgacagggtc 20
|||
DB 2372 GGCCTGCTGAACGGGATC 2391

RESULT 3

US-08-276-452A-72
Sequence 72, Application US/08276452A
Patent No. 5646029
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Allison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E

TITLE OF INVENTION: Plant Arabingalactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276.452A
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 85..480

FEATURE:
NAME/KEY: misc_feature
LOCATION: 85..147

OTHER INFORMATION: /note= "Putative secretion signal."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 154..393

OTHER INFORMATION: /note= "154-156,244-246,256-258,298-300,301-303,307-309 =
hydroxyproline."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 235..408

OTHER INFORMATION: /note= "235-268, 289-330, 331-365,
and 366-408 are internal peptide sequences from
amino acid sequencing."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 205..408

OTHER INFORMATION: /note= "235-268 = SEQ ID NO:67;
289-330 = SEQ ID NO:51; 319-357 = SEQ ID NO:68;
331-365 = SEQ ID NO:50; 366-408 = SEQ ID NO:53"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 154..393

OTHER INFORMATION: /note=
"310-312,325-327,337-339,361-363,378-380,391-393 =
hydroxyproline."
US-08-276-452A-72

Query Match 74.0%; Score 14.8; DB 1; Length 712;
Best Local Similarity 88.9%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gctgtctgacagggtc 20
|||

DB 412 GCTTCCTTGACAGAGTC 429

RESULT 4

US-08-798-744-72

Sequence 72, Application US/08798744

Patent No. 5830747

GENERAL INFORMATION:

APPLICANT: Chen, Chao-Guang

APPLICANT: Du, He

APPLICANT: Du, He

APPLICANT: Gane, Allison M

APPLICANT: Bacic, Antony

APPLICANT: Clarke, Adrienne E

TITLE OF INVENTION: Plant Arabidogalactan Protein (ACP) Genes

NUMBER OF SEQUENCES: 91

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee and Winner, P.C.

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: United States of America

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/798,744

FILING DATE: 13-FEB-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/276,452

FILING DATE: 18-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Caruthers, Jennie M.

REGISTRATION NUMBER: 34,464

REFERENCE/DOCKET NUMBER: 27-91A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)499-8080

TELEX: 49617824

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 712 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 85..480

FEATURE:

NAME/KEY: misc-feature

LOCATION: 85..147

OTHER INFORMATION: /note= "Putative secretion signal."

FEATURE:

NAME/KEY: misc-feature

LOCATION: 154..393

OTHER INFORMATION: /note=

OTHER INFORMATION: "154-156,244-246,256-258,298-300,301-303,307-309 =

OTHER INFORMATION: hydroxyproline."

FEATURE:

NAME/KEY: misc-feature

LOCATION: 235..408

OTHER INFORMATION: /note= "235-268, 289-330, 331-365,

OTHER INFORMATION: and 366-408 are internal peptide sequences from

OTHER INFORMATION: amino acid sequencing."

FEATURE:

NAME/KEY: misc-feature

LOCATION: 205..408

OTHER INFORMATION: /note= "235-268 = SEQ ID NO:67;

OTHER INFORMATION: 289-330 = SEQ ID NO:51; 319-357 = SEQ ID NO:68;

OTHER INFORMATION: 331-365 = SEQ ID NO:50; 366-408 = SEQ ID NO:53"

FEATURE:

NAME/KEY: misc-feature

LOCATION: 154..393

OTHER INFORMATION: /note=

OTHER INFORMATION: "310-312,325-327,337-339,361-363,378-380,391-393 =

OTHER INFORMATION: Hydroxyproline."

US-08-798-744-72

Query Match

Best Local Similarity 74.0%; Score 14.8; DB 2; Length 712;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gcttgctgaacaggtc 20

|||||

DB 412 GCTTCCTTGACAGAGTC 429

RESULT 5

US-08-975-762-43/C

Sequence 43, Application US/08975762

Patent No. 6207169

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975,762

FILING DATE: 21-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.439

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 3735 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-975-762-43

Query Match

Best Local Similarity 74.0%; Score 14.8; DB 4; Length 3735;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 gcttgctgaacaggt 19

|||||

DB 1618 GCTTCCTTGACAGAGTC 1601

RESULT 6

US-09-295-028-43/c
Sequence 43, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295.028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 43
LENGTH: 3735
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-43

Query Match 74.0%; Score 14.8; DB 4; Length 3735;
Best Local Similarity 88.9%; Pred. No. 65;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ggcctgttgacagagt 19
|||||
Db 1618 ggcttgcttgacagactgt 1601

RESULT 7
US-09-106-582-43/c
Sequence 43, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73 THERAPY
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106.582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 3735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-43

Query Match 74.0%; Score 14.8; DB 4; Length 3735;
Best Local Similarity 88.9%; Pred. No. 65;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ggcctgttgacagagt 19
|||||
Db 1618 ggcttgcttgacagactgt 1601

RESULT 8
US-08-932-376A-3
Sequence 3, Application US/08932376A
Patent No. 5869309
GENERAL INFORMATION:
APPLICANT: Pollino, Michael
APPLICANT: Tonzi, Sean M.
APPLICANT: Usher, John J.
APPLICANT: Burnett K, William V.
APPLICANT: Romanick, Guna
TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM
TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bristol-Myers Squibb Company
STREET: Rt. 206 & Provinceline Road
CITY: Princeton
STATE: New Jersey
COUNTRY: USA
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932.376A
FILING DATE: 17-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON0144a
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-4956
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2220 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-932-376A-3

Query Match 72.0%; Score 14.4; DB 2; Length 2220;
Best Local Similarity 93.8%; Pred. No. 96;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcctgttgacagagt 17
|||||
Db 680 ggcttgcttgacagag 695

RESULT 9
5248670-4/c
Patent No. 5248670
APPLICANT: DRAPER, KENNETH G.; ECKER, DAVID J.; MIRABELLI,
CHRISTOPHER K.; CROCKE, STANLEY T.
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR
INHIBITING HERPESVIRUS
NUMBER OF SEQUENCES: 15

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/485,297
FILING DATE: 26-FEB-1990
SEQ ID NO:4
LENGTH: 3688
5248670-4

Query Match 72.0%; Score 14.4; DB 6; Length 3688;
Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ctgcttgacagagt 19
|||||
Db 3248 CTTGCTTGACAGGTT 3233

RESULT 10
US-08-724-394A-20

; Sequence 20, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th floor
CITY: San Francisco
STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/724,394A

APPLICATION NUMBER: 01-OCT-1996

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..246240

OTHER INFORMATION: /note="HLA-H.CONFIG"

US-08-724-394A-20

Query Match 72.0%; Score 14.4; DB 2; Length 246240;
Best Local Similarity 93.8%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gggcttgctgacag 16
|||||
Db 58322 GGGCTGCTTGACAG 58337

RESULT 11
US-08-724-394A-21

; Sequence 21, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th floor
CITY: San Francisco
STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..246240

OTHER INFORMATION: /note="HLA-H.CONFIG"

US-08-724-394A-21

Query Match 72.0%; Score 14.4; DB 2; Length 246240;
Best Local Similarity 93.8%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gggcttgctgacag 16
|||||
Db 58322 GGGCTGCTTGACAG 58337

RESULT 12
US-08-724-394A-22

; Sequence 22, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.

APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note="HLA-H.CONTIG"
US-08-724-394A-22

Query Match 72.0%; Score 14.4; DB 2; Length 246240;
Best Local Similarity 93.8%; Pred. No. 2e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 15; Conservative 0;

Oy 1 gggctgtctgaacag 16
||||| |||||||

Db 58322 GGCGTGGCTTGACAG 58337

RESULT 13
US-09-039-555B-5
Sequence 5, Application US/0903955B
Patent No. 6033856
GENERAL INFORMATION:
APPLICANT: Koerner, Kathrin
APPLICANT: Mueller, Rolf
APPLICANT: Sadlacker, Hans-Harald
TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,555B
FILING DATE: 16-MAR-1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19710643.9
FILING DATE: 14-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 016779/0131
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Oligonucleotide"
US-09-039-555B-5

Query Match 71.0%; Score 14.2; DB 3; Length 79;
Best Local Similarity 84.2%; Pred. No. 67;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 ggcctgtctgaacagcgc 20
||||| |||||||

Db 53 GGCTTGCTGGACACAGTC 71

RESULT 14
US-08-473-020A-6/C
Sequence 6, Application US/08473020A
Patent No. 5877273
GENERAL INFORMATION:
APPLICANT: Hance, Allan J
APPLICANT: Grandchamp-Desreux, Bernard
APPLICANT: Levy-Freibault, Veronique
APPLICANT: Ciquel, Brigitte
TITLE OF INVENTION: Nucleotide sequences of actinomycetales,
TITLE OF INVENTION: applications to the synthesis or detection of nucleic
TITLE OF INVENTION: acids, products of expression of such sequences and
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Walter H. Dreger
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,020A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/623,729
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Walter H
REGISTRATION NUMBER: 24190
REFERENCE/DOCKET NUMBER: A54435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-473-020A-6

Query Match 71.0%; Score 14.2; DB 2; Length 343;
Best Local Similarity 84.2%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ggcttgctgaacaggtc 20
||| ||||| |||||
DB 211 GGCGGCTTGACGAGGCTC 193

RESULT 15

US-08-306-691B-46/C
Sequence 46, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
STREET: Two Penn Center, Suite 1800
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8363
TELEFAX: (215) 568-5549
TELEX: No. 5734039E
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 2638 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-306-691B-46

Query Match 71.0%; Score 14.2; DB 1; Length 2638;
Best Local Similarity 84.2%; Pred. No. 1.3e+02;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 ggcttgctgaacaggtc 19
||| ||||| |||||
DB 1703 GGCGGCTTGACGAGGCT 1685

Search completed: December 8, 2001, 10:53:12
Job time: 4982 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 09:33:20 ; Search time 288.04 Seconds
(without alignments)
59.528 Million cell updates/sec

Title: US-09-508-147-8

Perfect score: 20

Sequence: 1 ggcttgcctgacagagtc 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1661242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.GeneSeq_1101:*

1: /SIDS8/gcgdata/geneSeq/geneSeq/NA1980.DAT:*

2: /SIDS8/gcgdata/geneSeq/geneSeq/NA1981.DAT:*

3: /SIDS8/gcgdata/geneSeq/geneSeq/NA1982.DAT:*

4: /SIDS8/gcgdata/geneSeq/geneSeq/NA1983.DAT:*

5: /SIDS8/gcgdata/geneSeq/geneSeq/NA1984.DAT:*

6: /SIDS8/gcgdata/geneSeq/geneSeq/NA1985.DAT:*

7: /SIDS8/gcgdata/geneSeq/geneSeq/NA1986.DAT:*

8: /SIDS8/gcgdata/geneSeq/geneSeq/NA1987.DAT:*

9: /SIDS8/gcgdata/geneSeq/geneSeq/NA1988.DAT:*

10: /SIDS8/gcgdata/geneSeq/geneSeq/NA1989.DAT:*

11: /SIDS8/gcgdata/geneSeq/geneSeq/NA1990.DAT:*

12: /SIDS8/gcgdata/geneSeq/geneSeq/NA1991.DAT:*

13: /SIDS8/gcgdata/geneSeq/geneSeq/NA1992.DAT:*

14: /SIDS8/gcgdata/geneSeq/geneSeq/NA1993.DAT:*

15: /SIDS8/gcgdata/geneSeq/geneSeq/NA1994.DAT:*

16: /SIDS8/gcgdata/geneSeq/geneSeq/NA1995.DAT:*

17: /SIDS8/gcgdata/geneSeq/geneSeq/NA1996.DAT:*

18: /SIDS8/gcgdata/geneSeq/geneSeq/NA1997.DAT:*

19: /SIDS8/gcgdata/geneSeq/geneSeq/NA1998.DAT:*

20: /SIDS8/gcgdata/geneSeq/geneSeq/NA1999.DAT:*

21: /SIDS8/gcgdata/geneSeq/geneSeq/NA2000.DAT:*

22: /SIDS8/gcgdata/geneSeq/geneSeq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20	100.0	20	AAK75934
2	18	90.0	20	AAK75949
3	16.4	82.0	21	AAK75949
4	16	80.0	20	AAK75944
5	16	80.0	20	AAK75952
6	15.8	79.0	1102	AAK75952
7	15.8	79.0	1299	AAH34141
8	15.8	79.0	1652	AAH99221
9	15.2	76.0	318	AAK61694
10	15.2	76.0	381	AAK61693
11	15.2	76.0	980	AAK68115

12	15.2	76.0	1303	AAF26291	Pseudomonas sp Typ
13	15.2	76.0	2870	AAV71912	S. cerevisiae KIP2
14	15	75.0	20	AAV71912	Mouse oligonucleot
15	14.8	74.0	354	AAK60050	Human brain Expre
16	14.8	74.0	534	AAK14213	H. pylori GHP0 107
17	14.8	74.0	660	AAH53566	S. epidermidis ope
18	14.8	74.0	690	AAH07198	Human cDNA clone (
19	14.8	74.0	690	AAH22780	Human prostate can
20	14.8	74.0	698	AAH03283	Human cDNA clone (
21	14.8	74.0	712	AAO92528	N. alata arabinoga
22	14.8	74.0	903	AAK52837	Arabidopsis thalia
23	14.8	74.0	1102	AAK22774	Human prostate can
24	14.8	74.0	1559	AAH13671	Human cDNA sequenc
25	14.8	74.0	2348	AAK64683	cDNA encoding a hu
26	14.8	74.0	3504	AAH54495	S. epidermidis gen
27	14.8	74.0	3735	AAK6425	Ehrlichia sp. exte
28	14.8	74.0	3735	AAK6425	Human gene signalu
29	14.8	74.0	3735	AAK6425	Human gene signalu
30	14.4	72.0	20	AAK75953	Mouse oligonucleot
31	14.4	72.0	131	AAK26779	Human secreted pro
32	14.4	72.0	131	AAK26779	Human secreted pro
33	14.4	72.0	267	AAK68197	Human lung tumour
34	14.4	72.0	464	AAK16372	Human gene express
35	14.4	72.0	755	AAK33082	Human secreted pro
36	14.4	72.0	1940	AAH29852	S. cerevisiae apopt
37	14.4	72.0	2220	AAV21461	Rhodospiridium tor
38	14.4	72.0	2940	AAH46862	Human 25278 sulfat
39	14.4	72.0	20845	AAK6741	Human breast or ov
40	14.4	72.0	235033	AAV57926	Hereditary haemoch
41	14.4	72.0	117	AAV57903	Human secreted pro
42	14.2	71.0	117	AAK18283	Human secreted pro
43	14.2	71.0	343	AAK06509	Encodes protein XX
44	14.2	71.0	377	AAK6743	Novel human polynu
45	14.2	71.0	389	AAK17802	Human secreted pro

ALIGNMENTS

RESULT 1	
AAK75934	AAK75934 standard; DNA; 20 BP.
ID	AAK75934;
AC	AAK75934;
DT	29-JUL-1999 (first entry)
DE	Mouse p53 DNA fragment SEQ ID NO:8.
KW	Mouse; p53CP: p53; tumour suppressor; cell growth regulation;
KW	DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KW	tumour cell growth inhibition; genome guardian; differentiation;
KW	senescence; angiogenesis; ss.
OS	Mus musculus.
PN	W09925820-A1.
PD	27-MAY-1999.
PF	10-NOV-1998; 98WO-US23992.
PR	17-NOV-1997; 97US-0065740.
PA	(WARN) WARNER LAMBERT CO.
PI	Bian J, Sun Y;
PI	WPI. 1999-347468/29.
PT	New p53CP protein that specifically binds to the p53 consensus
PT	binding sites, useful for treating p53 associated disorders

good date
Appd.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147199.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 82.0%; Score 16.4; DB 21; Length 1281;
Best Local Similarity 94.4%; Pred. NO. 48;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ggcctgcctgaacagcgt 19
|||||
Db 1125 ggcctgcctgaacatgct 1142

RESULT 4
AAx75944/c
ID AAx75944 standard; DNA: 20 BP.

XX AAX75944;
AC 29-JUL-1999 (first entry)
DT
XX
XX
DE Mouse oligonucleotide T3SF SEQ ID NO:9.
XX
XX Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KM DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KM tumour cell growth inhibition; genome guardian; differentiation;
XX senescence; angiogenesis; ss.
OS Mus musculus.
XX
XX WO9925820-A1.
PN
XX 27-MAY-1999.
PD
XX 10-NOV-1998; 98WO-US23992.
XX
XX 17-NOV-1997; 97US-0065740.
PR
XX (WARN) WARNER LAMBERT CO.
PA
XX Bian J, Sun Y;
PI
XX WPI; 1999-347468/29.
DR
XX New p53CP protein that specifically binds to the p53 consensus
PT binding sites, useful for treating p53 associated disorders
XX
XX Disclosure; Page 7; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kD) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
XX
XX Sequence 20 BP; 5 A; 8 C; 4 G; 3 T; 0 other;
SO

Query Match 80.0%; Score 16; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ttgcttgacacaggtc 20
|||
Db 20 TTGCTTGACACAGGTC 5

RESULT 5
AAX75952
ID AAX75952 standard; DNA; 20 BP.
XX
XX AAX75952;
AC
XX 29-JUL-1999 (first entry)
DT
XX Mouse oligonucleotide T3SF4.
DE
XX Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KM DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KM tumour cell growth inhibition; genome guardian; differentiation;
XX senescence; angiogenesis; ss.
XX Mus musculus.
XX
XX WO9925820-A1.
PN
XX 27-MAY-1999.
PD
XX

PF 10-NOV-1998; 98WO-US23992.
XX
XX 17-NOV-1997; 97US-0065740.
PR
XX (WARN) WARNER LAMBERT CO.
PA
XX Bian J, Sun Y;
PI
XX WPI; 1999-347468/29.
DR
XX New p53CP protein that specifically binds to the p53 consensus
PT binding sites, useful for treating p53 associated disorders
XX
XX Disclosure; Page 22; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kD) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
XX
XX Sequence 20 BP; 3 A; 3 C; 5 G; 9 T; 0 other;
SO

Query Match 80.0%; Score 16; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 ctgcttgacacaggt 19
|||
Db 3 ctgcttgacacaggt 18

RESULT 6
AAF18281/C
ID AAF18281 standard; DNA; 1102 BP.
XX
XX AAF18281;
AC
XX 14-MAR-2001 (first entry)
DT
XX Lung cancer associated polynucleotide sequence SEQ ID 300.
DE
XX Human; lung cancer associated protein; neuroprotective; cytoskeletal;
KM cardiocative; immunomodulatory; muscular active; vunerary;
KM gastrointestinal; nephrotropic; antinfetive; gynecological;
KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KM proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
OS
XX WO20005180-A2.
PN
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05918.
PF
XX 12-MAR-1999; 99US-0124270.
PR
XX Ruben SM;
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX WPI; 2000-587514/55.
DR P-PSDB; AAB58405.
DR
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX Claim 1; Page 758-759; 1425pp; English.
PS

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardiocactive;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrotoxic; antinefactive; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1102 BP: 249 A; 328 C; 284 G; 237 T; 4 other.

Query Match 79.0%; Score 15.8; DB 21; Length 1102;
Best Local Similarity 89.5%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcgtgcttgacagagtc 20
||| |
Db 86 GGCGTTCTTGAAACAGGTC 68

RESULT 7
AAH34141/c
ID AAH34141 standard; CDNA: 1299 BP.
XX
AC AAH34141;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding CDNA SEQ ID NO:1223.
XX
DE Human colon cancer antigen encoding CDNA SEQ ID NO:1223.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PM WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HOMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
XX
P-PSDB: AAG74736.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3000; 9803pp: English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 1299 BP: 294 A; 385 C; 347 G; 269 T; 4 other.

Query Match 79.0%; Score 15.8; DB 22; Length 1299;
Best Local Similarity 89.5%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcgtgcttgacagagtc 20
||| |
Db 283 GGCGTTCTTGAAACAGGTC 265

RESULT 8
AAH99221/c
ID AAH99221 standard; CDNA: 1652 BP.
XX
AC AAH99221;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding CDNA sequence SEQ ID NO:56.
XX
DE Human protein encoding CDNA sequence SEQ ID NO:56.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; vitruide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antiaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
PM WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-457603/49.
XX
P-PSDB: AAM25280.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the

PF treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1; Page 325; 1217pp; English.
XX
CC AA99166 to AA99904 encode the human proteins given in AAM25225 to
CC AAM25993. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antihistaminic; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocidic; anti-HIV; fungicidal; antitumour;
CC cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
CC anticancer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cyostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 1652 BP; 323 A; 516 C; 482 G; 331 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 1652;
Best Local Similarity 89.5%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgctgaacaggc 20
||| ||||| ||||| |||||
Db 694 GGCCTTCTTGACAGCGTC 676

RESULT 9
AA61694/c
ID AAX61694 standard; DNA; 318 BP.
XX
AC AAX61694;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, t805.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PI 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI: 1999-189980/16.
DR P-PSDB; AAY19997.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases

Hand date

PF caused by Borrelia, particularly Lyme disease
XX
PS Claim 1; Page 155; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 318 BP; 126 A; 49 C; 51 G; 92 T; 0 other;

Query Match 76.0%; Score 15.2; DB 20; Length 318;
Best Local Similarity 85.0%; Pred. No. 1.7e+02; 3; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcctgctgaacaggc 20
||| ||||| ||||| |||||
Db 176 GAGCTTGCTTGACAGCATC 157

RESULT 10
AA61693/c
ID AAX61693 standard; DNA; 381 BP.
XX
AC AAX61693;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, f805.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PI 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI: 1999-189980/16.
DR P-PSDB; AAY19996.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
SQ Claim 1; Page 155; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 381 BP; 150 A; 59 C; 62 G; 110 T; 0 other;


```
DE 5. cerevisiae KIP2 DNA sequence.
XX
XX TIH1; recombinant; research; epitope mapping; modulating; CKI; Yeast;
KM casein kinase I; cell growth; KIP1; cancer; viral infection; ss.
XX
XX Saccharomyces cerevisiae.
OS
XX US5846764-A.
XX
XX 08-DEC-1998.
XX
XX 23-JAN-1995; 95US-0376843.
XX
XX 23-JAN-1995; 95US-0376843.
PR 21-JAN-1994; 94US-0184605.
XX
XX (ICOS-) ICOS CORP.
PA
XX Demaggio AJ, Hoekstra MF;
PI
XX WPI; 1999-059057/05.
XX
XX Polynucleotide encoding yeast TIH1 polypeptide - useful for
PT producing recombinant polypeptide
XX
XX Example 3; Columns 235-238 (columns 61-64); 46pp; English.
XX
XX This represents the DNA sequence of Saccharomyces cerevisiae KIP2 that
CC belongs to the yeast kinesin gene family. The invention relates to TIH
CC proteins that interact with casein kinase I (CKI) isoforms. A host cell
CC transformed with a DNA expression construct containing the TIH1 DNA can
CC be used to produce recombinant TIH1 polypeptide which is used for
CC research purposes. The TIH proteins are used to map TIH epitopes in the
CC development of epitope-specific agents that may be useful for modulating
CC CKI/TIH protein interactions involved in cell growth in health and
CC disease, e.g. cancer or viral infections.
CC Note: The specification has erroneous column number indications.
XX
XX Sequence 2870 BP; 964 A; 559 C; 615 G; 732 T; 0 other;
XX
XX
XX Query Match 76.0%; Score 15.2; DB 20; Length 2870;
XX Best Local Similarity 85.0%; Pred. No. 2.2e+02;
XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 ggagcttgctgaacagggtc 20
XX |||||||||
DB 2372 ggcttgctgaacggggtc 2391

RESULT 14
AA75951
ID AAX75951 standard; DNA; 20 BP.
XX
XX AAX75951;
AC
XX 29-JUL-1999 (first entry)
DT
XX
XX Mouse oligonucleotide T3SF3.
DE
XX
XX Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KM DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KW tumour cell growth inhibition; genome guardian; differentiation;
XX senescence; angiogenesis; ss.
XX
XX Mus musculus.
OS
XX
XX WO9925820-A1.
XX
XX 27-MAY-1999.
XX
XX 10-NOV-1998; 98WO-US23992.
XX
```

```
PR 17-NOV-1997; 97US-0065740.
XX
XX (WARN ) WARNER LAMBERT CO.
PA
XX
XX Bian J, Sun Y;
PI
XX
XX WPI; 1999-347468/29.
XX
XX New p53CP protein that specifically binds to the p53 consensus
PT binding sites, useful for treating p53 associated disorders
XX
XX Disclosure; Page 22; 37pp; English.
XX
XX The present invention describes a p53 competing protein designated p53CP
CC (40 kd) that specifically binds to the p53 consensus binding sites.
CC The p53CP protein is useful for p53 inactivation and regulation during
CC human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
CC growth inhibition, genome guardian, differentiation, senescence, and
CC angiogenesis.
CC
XX
XX Sequence 20 BP; 3 A; 2 C; 5 G; 10 T; 0 other;
XX
XX
XX Query Match 75.0%; Score 15; DB 20; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ttgcttgacagggt 19
XX |||||||
DB 3 ttgcttgacagggt 17

RESULT 15
AA060050/C
ID AA060050 standard; DNA; 354 BP.
XX
XX AA060050;
AC
XX
XX 16-MAR-1994 (first entry)
DT
XX
XX Human brain Expressed Sequence Tag EST02032.
DE
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
KM transcription; mapping; locations; chromosomes; Chromosomal; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9316178-A.
XX
XX 19-AUG-1993.
XX
XX 12-FEB-1993; 93WO-US01294.
XX
XX 12-FEB-1992; 92US-0837195.
PR
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PA
XX
XX Adams MD, Moreno RF, Venter CJ;
PI
XX
XX WPI; 1993-272882/34.
XX
XX
XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
XX
XX Example 4; Page 273; 500pp; English.
XX
XX The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
```

CC type, and for prepn. of antisense sequences, probes and constructs.
 CC EST02032 has a "excellent" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also AA059041-061440.
 XX
 SQ Sequence 354 BP; 93 A; 104 C; 99 G; 54 T; 4 other;

Query Match 74.0%; Score 14.8; DB 14; Length 354;
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 gggcttgcttgacagggtc 20
 |||| ||||| | |||| |
 DB 298 GGGCTGCTTGNANAGGGGC 279

Search completed: December 8, 2001, 12:14:07
 Job time: 9647 sec

THIS PAGE BLANK (USPTO)

10/11/98

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 09:28:55 ; Search time 1723.37 Seconds

(without alignments)
191.453 Million cell updates/sec

Title: US-09-508-147-8

Perfect score: 20

Sequence: 1 gggcttgcctgaacagggc 20

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: GenBank:
2: gb_ba:*
3: gb_hg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.4	87.0	13245	1	AE001977	AE001977 Deinococcus
2	17.4	87.0	296820	10	AF312033	AF312033 Mus muscu
3	16.8	84.0	1950	9	AK026710	AK026710 Homo sapi
4	16.8	84.0	2337	10	AF019974	AF019974 Rattus no
5	16.8	84.0	11021	1	AE004674	AE004674 Pseudomon
6	16.8	84.0	156640	2	AC013777	AC013777 Homo sapi
7	16.8	84.0	182110	2	AC027466	AC027466 Homo sapi
8	16.8	84.0	184043	2	AC068806	AC068806 Mus muscu
9	16.4	82.0	961	2	AC063109	AC063109 Giardia i
10	16.4	82.0	968	2	AC052567	AC052567 Giardia i
11	16.4	82.0	3010	1	AF116285	AF116285 Pseudomon
12	16.4	82.0	12357	1	AE004471	AE004471 Pseudomon
13	16.4	82.0	83499	8	AT11008	AT11008 Arabidops
14	16.4	82.0	101715	8	ATF4D11	ATF4D11 Arabidops
15	16.4	82.0	108387	8	AC016662	AC016662 Arabidops
16	16.4	82.0	118507	8	ATF7J8	ATF7J8 Arabidops
17	16.4	82.0	135268	2	AC016937	AC016937 Homo sapi
18	16.4	82.0	141710	2	AC012348	AC012348 Homo sapi
19	16.4	82.0	147552	2	AC019298	AC019298 Homo sapi
20	16.4	82.0	151329	2	AC025623	AC025623 Homo sapi
21	16.4	82.0	158628	2	AC020653	AC020653 Homo sapi
22	16.4	82.0	158789	2	AP003398	AP003398 Homo sapi
23	16.4	82.0	174129	2	AC012236	AC012236 Homo sapi
24	16.4	82.0	176330	2	AC078835	AC078835 Homo sapi
25	16.4	82.0	197252	8	ATCHR1V77	ATCHR1V77 Arabidops
26	16.4	80.0	2615	8	YSCA4P1	YSCA4P1 Arabidops
27	16.4	80.0	3910	8	YSCA4P2	YSCA4P2 Arabidops
28	16.4	80.0	41200	8	SC8520X	SC8520X Saccharomyc
29	16.4	80.0	44113	8	YSC84179	YSC84179 Arabidops
30	16.4	80.0	164879	9	AC007876	AC007876 Homo sapi
31	16.4	80.0	166955	2	AC080161	AC080161 Homo sapi
32	16.4	80.0	199359	2	AC074027	AC074027 Mus muscu
33	16.4	80.0	216456	2	AC026227	AC026227 Homo sapi
34	16.4	80.0	239574	2	AC073791	AC073791 Homo sapi
35	15.8	79.0	360	1	AF241205	AF241205 Mycobacte
36	15.8	79.0	7292	1	HSCST3G	HSCST3G Human GST3
37	15.8	79.0	10677	1	AE005967	AE005967 Caulobact
38	15.8	79.0	14006	1	AE002049	AE002049 Deinococc
39	15.8	79.0	23143	8	FI5P11	FI5P11 Arabidops
40	15.8	79.0	61921	9	HSDJ892M9	HSDJ892M9 Human DNA
41	15.8	79.0	73131	2	AC016443	AC016443 Homo sapi
42	15.8	79.0	110000	2	AL591074	AL591074 Homo sapi
43	15.8	79.0	120766	2	AC004150	AC004150 Homo sapi
44	15.8	79.0	128822	2	AP001260	AP001260 Homo sapi
45	15.8	79.0	130049	8	NC17E5	NC17E5 Neurospor

ALIGNMENTS

No data

RESULT 1
AE001977/c
LOCUS 13245 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans RI section 114 of 229 of the complete
ACCESSION AE001977 AE000513
VERSION 1
KEYWORDS Chromosome 1.
SOURCE
ORGANISM Deinococcus radiodurans.
Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE
1 (bases 1 to 13245)
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.R., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Ulteback, T., Zaleski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
JOURNAL Science 286 (5444), 1571-1577 (1999)

MEDLINE 20036896
REFERENCE 2 (bases 1 to 13245)
AUTHORS White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
Dodson,R.J., Haft,D.H., Gwin,M.L., Nelson,W.C., Richardson,D.L.,
Mofat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
Vanatchevan,J.J., Lam,P., McDonald,L., Uterback,T., Zaleski,C.,
Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE Direct Submission
JOURNALS Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1. 13245
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
94. .906
/gene="DR1296"
94. .906
/note="similar to PID:349284 percent identity: 55.65;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="transposase, putative"
/protein_id="AAFI0868.1"
/db_xref="GI:6459046"
/translation="MSEKPKYKSDDDERTYFLPFLAPEDAHORVPIREVLNA
ALMIGRTGSCWEYLPHPFPKTYHQOLMRFEFGCEENLAHDSLVREDALEGPV
TVAIVDSRTLOSTPESGRAGYDGGKRRKSKIAADVDMGNVMTLLVTPGNECDRQ
VYDCREVQOVTGDHIDVIVADQGEQPIDASIDVELVVKRPTGATGEVLLPL
RMVVERFPAMTARFRRRLSRDLERLQSSLGFGHMLAVSVTLNKLKPLTGLSA"
988. .1923
/gene="DR1297"
988. .1923
/note="similar to GB:A1123456 percent identity: 72.48;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAFI0869.1"
/db_xref="GI:6459047"
/translation="MSGGLVALDDVAIAIKLAASVYDICAAGRAGAKAVYVVD
TATPRVYAGRPRLPYIMKIAKSLFNTIVILPVIILISQFVPMANPLMIGG
LYTSFEGAEKIEALITGSDHDEADAAVGTPEHEKOMSGAITDITLSLTIYAL
AEYTEQALFRAVTLVIALITLTLVGVGLVAKMDIGLMAQGRIGTTRALGRGL
VRGMPVMSVLSVVGTAAMLVGGHILVSGLEKFGMLAHAIYHDAEAGHALPLFG
GAVTWETLGSAAVGLAVGALIVGLHLPRKEH"
2108. .3676
/gene="DR1298"
2108. .3676
/note="similar to PID:1653565 percent identity: 59.91;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAFI0870.1"
/db_xref="GI:6459048"
/translation="MTEALQPLRLISYWNKEIKPLDAETGTLFKQADIRVLAEP
NRYSVGMAISGLQYIYRMFNOEGVACERAFLPDVPDAPERTGALPTVSGRAAGC
ELFSISVSEFLDNLNITRLIDVAGMPLAREBDETDVVMIGPILSSNYPPLTPFD
IITIGDEQIVPVVSEALRESSREDYDILIDGYPGFPLARHYHEPTATAREKLLP
AISQIVTPHSELSELMFLVEAQRGCPRCTPLCATMTGPNKNNQAGLLDVIIPDWK
VGVGALISPEPHTKFVGKTLTERGVKLGVSIRADVDAELAILKAGLIRFTVS
DAESERLRMLKGIITEDLKTQAQISRDLSGSLKYMMIGLGPENDDITELISPT
KELAGINRIALGISPEVKRHTPHFADPFGVGTIEKRLRIKELRTTAEIRNVSAS
MAWVESYIARGGPEVGMAYQIYRNESIGAKKALDEVGMDDFEANTPSIGLPGQY
ESKEVSAAHOGGLAY"
complement(3746. .3967)

CDS
/gene="DR1299"
complement(3746. .3967)
/gene="DR1299"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAFI0878.1"
/db_xref="GI:6459056"
/translation="MAQGVMTNLSWTPTQKRRRLATLIAAYLATSAGLIAAGLHMP
WYLAISLVLYGVLYVGYAVLKAAYRA"
complement(3952. .4644)
/gene="DR1300"
complement(3952. .4644)
/note="similar to GB:A1123456 percent identity: 61.57;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="cytochrome c-type biogenesis protein, putative"
/protein_id="AAFI0871.1"
/db_xref="GI:6459049"
/translation="MIPGVNSAPSLTYAFLAGLISFSLPCVPLIPSLYGLGNAK
APLGRALGTAGGCLVFAUGATSTGATLAPKLLGOLAAYLVIFELVMDGVR
LPFLMRDTROLADGGYGPVALGAFAFGWSPCLGSLGILAASSASLGSVRL
VAYNLGLAPFLAALLMDRLNRLRNRVAGIEFEKVGAVLVGVMMLTGQFTRLAT
FFFSMPMEWELK"
complement(4673. .5368)
/gene="DR1301"
complement(4673. .5368)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAFI0879.1"
/db_xref="GI:6459057"
/translation="MPVHLKTVFFSTAPQKLRPDMGSLNLOPYOLVGRMPPPGV
WSRNSMEDITMSGFPAIVEGFPLVREKHHHOMRLHEERALQAAQAPVPGV
DAPRALIRPEPHLYLVALLCRLQEDAPLAGDARARYLISOARGELPTTLRAYNL
TPARQOLTNQGNPPELLLEQLEMSQGYEALGRSAAADRLMAGHILROKPOPI
ELGEPVAERKRS"
5550. .6638
/gene="DR1302"
5550. .6638
/note="similar to GB:M64519 SP:P23858 PID:147326 GB:U00096
PID:1651555 percent identity: 64.06; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="spermidine/putrescine ABC transporter,
ATP-binding protein"
/protein_id="AAFI0872.1"
/db_xref="GI:6459050"
/translation="MSVVDVFKSYRNAAGETPYLKDIDLRIGEFFSLGSGCGK
TTLIRIAGFEQADSGAVVIGGRDMGIPAHRRPVNVFQSYALFPLSVADWAVGL
KOKGVRAELRORVERAMETVOIGOFENRPPDLSGROVALARIVAGPVLID
EPLSALDLIKRKOLOVELSNLOEDLGTTFVFAHDEQEAIVMSDRIAVNMQGRLEQD
RAPDIYRPTATAPANTLSSNLNLAGVALHSTEARQVEPGRPLTHRAGISYQD
VTLSIRPEKLRMERDDETEGNEVRAVDIVYGAENQYLLEAHGRLVAFQUNADIG
ADDDFDYDEEVALYLPASLVLEEBA"
6778. .7596
/gene="DR1303"
6778. .7596
/note="similar to GB:M64519 SP:P23860 PID:147327 GB:U00096
PID:1651552 percent identity: 70.92; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="spermidine/putrescine ABC transporter, permease
protein"

gene

[illegible]

NC_147615.1:MSGSEFMDLILDDIKADAIYVMGLAAVYLKMGEGFNDRITILEOPEEES
AGTKEAEAKSKTEARAPALGEGERKANDDKERKKDGAOLEHSSNDNRTKTKSBSDGR
EEKTEAEKEAKSKRNKRQSGDSDEGSVESLESBGQAEEKEREALEALKER
EKPKPEERKPDAGLECKRPLPHKTSJEMNIANIISAEIITSICRYPGPMVA
LSBPPEERFRFGWTFEDRSVNKEICMIONLRRECELSPCVNDLTRVINING
ITOKOIYRNDIKTLAATLIHTLDROTROSLANSEPTPVPSITONLIKNTIDYLE
EVSAEPEELIGSSGGPPEEPKRGNAELINVEDEKLIVLKILLIYLVHSIDXY
NTCEYPNDEPRNRGCIIHVNGPPMRPNRISHGEVLNMOKTPEEKUPLLVSRESLEE
EAOAGRRDPDEQEVEKEVTSTOLGKDKNLCPLSGKFKPGFEVRNHINKHAETIE
EAKKEVAFENNELDKARPALPEIKPPAQPPPAIILPPGLTPGIAPHQPOGLAMPYG
QPRLPIGYGVTPPYPHAPYGAGGNDAFQGCGCYGPKPRNRVARDPRAIVEY
RDLPADVDVDF"

complement(join15)098...15173,15254...15359,15485...15574,
15734...15890,16042...16235,16346...16499,16607...16780,
17066...17254,17495...17573,17658...17719,17993...18150,
18825...18931,19021...19119,19234...19418,19511...19680,
21333...21581,21795...21941,22180...22308,26327...26448)
/gene="Asr2"
/codon_start=1
/evidence-not_experimental
/product="Asr2c"
/protein_id="AAK28820.1"
/db_xref="GI:13517495"
/translation="MGSDSYDERRRRDKFRERSDYDRSREDRRRRGDMWDREMD
RGRRRSRGERTDYDRNRREFSPRIEISPPQKRMKMDHESSDYSHGCTMPYVG
GGGGPTGPPQPMGWPHVIHQHVLFTQARLGSIAIDIVIPYIKSTRKEFLSLD
DSVDTEAVKRYNDYKLDIFRQQQODEFLKADEWRSVYHPDEPKRKOEALGLQ
NRLVFLSLMSGWFMDLLDIDKADIIVMLAAVYLMKGEGTENDIRILEOPEEEO
AGTKEAEAKSKTEARAPALGEGERKANDDKERKKDGAOLEHSSNDNRTKTKSBSDGR

Query Match 87.0%; Score 17.4; DB 10; Length 296820;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ggccttcggaacaggtc 20
|||
Db 216826 GCTTGCTTGACAGAGC 216844

	RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT			
	AK026710/c	AK026710	Homo sapiens cDNA: FLJ23057 fls, clone LNC03755.	AK026710.1	GI:10439626	oligo capping: fls (full insert sequence).	Homo sapiens human lung cDNA to mRNA, clone_lhb clone:LNC03755.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	Kawakami,T., Noguuchi,S., Itoh,T., Shigeta,K., Senba,T., Matsunura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M., Omori,Y., Ota,T., Suzuki,Y., Oiyashiki,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T., and Sugano,S.	NEDO human cDNA sequencing project	Unpublished (2000)	2 (bases 1 to 1950)	Sugano,S., Suzuki,Y., Ota,T., Oiyashiki,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T., and Nakamura,Y.	Direct Submission Submitted (29-AUG-2000) to the DDBJ/EMBL/genbank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology Agency).

FEATURES
SOURCE
1. 1950
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LNC03755"
/clone_lib="LNC"
/tissue_type="human lung"
/note="Cloning vector pME18SF13"

BASE COUNT
576 a 402 c 446 g 526 t

ORIGIN

Query Match
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 99gctgctgaacaggctc 20
||||| ||||||| |||
Db 390 GGGCTGGCTTGACAGAGTC 371

RESULT 4
AF019974 2337 bp mRNA ROD 04-OCT-1997
DEFINITION Rattus norvegicus chromogranin B (Chgb) mRNA, complete cds.
ACCESSION AF019974
VERSION AF019974.1 GI:2465397
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 2337)
Forss-Petter,S., Danielson,P., Battenberg,E., Bloom,F. and
Sutcliffe,J.G.
Nucleotide sequence and cellular distribution of rat chromogranin B
(secretogranin I) mRNA in the neuroendocrine system
J. Mol. Neurosci. 1 (2), 63-75 (1989)
90351888
2 (bases 1 to 2337)
Forss-Petter,S., Danielson,P.E. and Sutcliffe,J.G.
Direct Submission
Submitted (18-AUG-1997) Molecular Biology, Scripps Research
Institute, 10550 North Torrey Pines, La Jolla, CA 92037
Location/Qualifiers
1. 2337
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. 2337
/gene="Chgb"
64. 2091
/note="secretogranin I"
/codon_start=1
/product="Chromogranin B"
/protein_id="AAB72089.1"
/db_xref="GI:2465398"
/translation="MORALVIGLGAALAAVISAAPVNDPNDNEEYTRCIIEYLNSA
LSKSAPIITPEECROVLRLKSGEYKGEKGENSKFEVRLLRPSDSVSRWASSRE
ETGAPVEDSPQAAVNDNEWTGGGSHREAVDDDSLAPSNQVSKKAIHSEBREG
KEEDEBEKTIYPKEHNRDAGEKHTHEESGEKHNASNKRSASAKKESVARAE
HFELEKTHSREDSQESGEETROEKQELPDODQSESEEGEGEGCATSVYTKR
RPRHHHNSQNSKSYEGRGLSEERKHAAGESDANAYATNLGKRGHHLAHYRASE
EEDYGEELRSYGPFOAGLOVRGSGSEVRAPSRSESEKRYKRNHPSLST
ANRSEETEERSTEGAKRGHRGRGPRPAYPALDSRQERLDEGHDPVHESPVDT
AKRYQSKWMOQGQENLYNLTDESGDQGRWMOQEOLEPESREEVTSFPRQAPYPTTE
KRRRLGALFNYPDPLQWKNSDPERKGNPDSDSLDDGEGDNGVTMTENKFNPFYND
IMWKRPFSEIDNMVYKRSFAKAPHLDLKROYDGVAVELDQLLHYRKKAAFPPEYDS
EEQGGPHQEADEKDRADQRYLTPEEKKELLENLAAMDELOKIAEKTSQR"

BASE COUNT
753 a 524 c 703 g 357 t

ORIGIN

Query Match
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 99gctgctgaacaggctc 20
||||| ||||||| |||
Db 1403 GGGCTTGATGACAGGGTC 1384

RESULT 5
AE004674 11021 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 235 of 529 of the complete
genome.
ACCESSION AE004674 AE004091
VERSION AE004674.1 GI:9948512
KEYWORDS
SOURCE Pseudomonas aeruginosa.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 11021)
Stover,C.K., Pham,X.-Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Huftagle,W.O., Kowalik,D.J., Lagrou,M.,
Gardner,R.L., Goltz,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 11021)
Hickey,M.J., Brinkman,F.S., Huftagle,W.O., Kowalik,D.J.,
Lagrou,M., Gardner,R.L., Goltz,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1. 11021
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
/complement(87. 1784)
/gene="PA2463"
/complement(87. 1784)
/gene="PA2463"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05851.1"
/db_xref="GI:9948513"
/translation="MRSYRAYVAVGIVLPLASAPFTSPGRDLTRDQORLLDE
ORRRLEQQLPGGAPAAVAGSGDDCEFRIRLEAGHLEGSARQULAPYQGR
CLGVQELNALKAVTDHLDLGRGVYTRAYLQODLASGTLPIVEGRLGIDSSALA
SPRELAMSFPERGTGELDLRLGQVLDLSRPSQADLELVPSGSEVGSRYRKGSR
DKPWRVATRRNDDGVSTGEOMGIGLDMDSPLGLADLNRANDAVTDHWRHSDQ
SLFSLPMGWMTFTYGVSDYRTFRNMSGPFLDGDSRSRHPAEAVNLDFGMOGAL
AMSGLSHQRRNNYLDPRLDQSTRTETQGLNHRGRTSGCVNADLDFGMOGAL
GAQGGHHPHADPNARIDKYSLTSLYDLPFLWGBRSFSDSLATGQREEDVLFSPQRI
SLGNSSVVGKRDQTLTGDSGGVWNRQLRMRRAVAVAPLWLEYGVAFAVAVGVIR
HDRVNDGASGMRGNAIELDARGRFASVGFARSLSRPSAIERREHPIYRVDAFF"
/complement(2097. 2621)
/gene="PA2464"

```
CDS
    complement(2097..2621)
    /gene="PA2464"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAG05852.1"
    /db_xref="GI:9948514"
    /translation="MNAIRIALGLLLPLAAGSAVAAREGSSVKKDAKAASAIETGK
    NDLGVSBSGITSGRGSAOGADAKVSDNAGFAEIGVEVLTKEPRDQQLVTLTGFK
    GDATVRLINLTORGALLVIDNDGYSNALVALANPDDVTYPAKAGIRQTFVEEGAGS
    IKAVRMKCDLPRK"
    complement(2668..3816)
    /gene="PA2465"
    complement(2668..3816)
    /gene="PA2465"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAG05853.1"
    /db_xref="GI:9948515"
    /translation="MRPVLLHRYVGLATALLPLFLAGLTGSLAFHHEIDEMLPGF
    YAVGGERLSPGLVORVESRYPOLVWMEYEPAGHPALLATVPREAGAKKEHDV
    FYLDVSGEEYKRIAMACCPANLIVPWLEPHHNLTLPGNMGLYLMGAMFWLID
    CFVGAMLLTPRGRPPMSKWTYAMIKRGNAYRFNFDLHRAGGLMLMLLAVALSVA
    LNLPSQVFKPLVLSLSPLESVYDARGRLPREQGETRDLDTFOLASVBAARLGLA
    EPIGELYSEFENFAGFGDHDPMGKSMLEFFHSGDRLDGOEAVAGSGWGERFYRL
    QYPIHGRIAGLPRIAIAGLISLTGVYIMWKKRRARHMNGR"
    complement(3934..6396)
    /gene="PA2466"
    complement(3934..6396)
    /gene="PA2466"
    /codon_start=1
    /transl_table=11
    /product="probable TonB-dependent receptor"
    /protein_id="AAG05854.1"
    /db_xref="GI:9948516"
    /translation="MTATAVVLRNAPSLSDFPPARSRLSRVRAALLSLAMAAGAPLC
    ASAAEAEOARPYAIPAGOLGVINREARAGITLSTPATGTGSSOGKSGFTVQ
    OGARLADPLLEADOGDSFVLEAPAKGDVILNQAVFALGNNGLSTGCVLT
    HSOIATKSKRLTSTQVSYVITREQIDDPRTASKVQCAMRTPTGIFTCQVASRYX
    VVMGFADNSVDNITLDGLKAMGDSGTSSMQVDPYFLERIDVLKGPSSVLYGSLPG
    GLVALTSKRLYEDYKQITGSGMGEKMFDSGLDEKRLIARLYLGKSGSDQ
    FDHYKERYAIAPLALIDSDDTLTLLQGYLDHPNGRYHGGVPADGTLSHHNGRHS
    REFDGSPSKDDEPRTQMEFYOLEHRIIDVWSARONFRYLDSDVDSQVAYAWSAS
    EPNKLNRYFSAGREHLIAYIDNMLQAEFAFGARHTLLGLDQRRTYVDMRSGSA
    SALDAFNPVYGDAISYFPDNNHTRLEQNGVYODLIDIDOMRESGLRDQWVSVD
    KNRSTGSKADDDMEKFTGRIGALYLPNGLAPYISYSEFPNAYSDASGTPPLAPTES
    KQWELGLFOAPGNSNFYASLFIHTDENASKPQDNFTYSVEGVSQGLSELAHTQ
    LSDMLKLGSYTYDITYTKSLDNGOHTPNQARKHMASLMADYAFDAGPLSGLSIG
    GARVYGTWADKENTLAVPDYTLVDARIGYDLGKLGKLDVSLANNLLDKDVASC
    YSLDFCYEGEKRRNTATVNYOF"
    complement(6540..7526)
    /gene="PA2467"
    complement(6540..7526)
    /gene="PA2467"
    /codon_start=1
    /transl_table=11
    /product="probable transmembrane sensor"
    /protein_id="AAG05855.1"
    /db_xref="GI:9948517"
    /translation="MDGTRGVDEAVVROAILHMLVRLRSQPADRLQACAMRAEHG
    EHERAQQOVSALNELQGRFALPGVAYGLSDSARQALKLSLLVAGGAVA
    WNGRDSLPQRLSADYSTATGERSIELVQGTROLQMTDSADVDRYAGOGILLAG
    EPLASGADNOSPTRHRLVTRTAGLEALGTRNNVRLQDQATLSVBSGVSRLDARG
    ARPLQAPVALEGOSYRTIADGVRRLERPDQADAMAGLLVTRDMRLADFLAEVARR
    NGYLGCAAEVADRLSGVYRLDDTKLLQVLAARLPVLRHRTRMWVYVAA"
    complement(7535..8053)
    /gene="PA2468"
    complement(7535..8053)
    /gene="PA2468"
    /codon_start=1
    /transl_table=11
    /product="probable sigma-70 factor, ECF subfamily"
```

```

    /protein_id="AAG05856.1"
    /db_xref="GI:9948518"
    /translation="MSAPILSARHRAHLYSEHHGWLQNMRLGKLGCAADADLQ
    DTFRLILKRELREIGMRAPLRTIARGVLDHRRRETORAYLESTIAHLEQAPSP
    EARELVLELEIEISRMIDGLPKVRYATFLAACEDLSHROLAERMGSQSRVEYVAE
    ALHYCYLIRIGE"
    complement(8142..9056)
    /gene="PA2469"
    complement(8142..9056)
    /gene="PA2469"
    /codon_start=1
    /transl_table=11
    /product="probable transcriptional regulator"
    /protein_id="AAG05857.1"
    /db_xref="GI:9948519"
    /translation="MDMTRRLRLQLOLVSLMETGNLSTGARAHTTPOPLSKMLE
    LEEDVAPLPERHARGRLAPRYGMLTFPHNARVLTETEMARONLEAMRSGSGSVLLG
    TSPASPSLVRAIRLARGITPGQVVELLEBTNGLELERKROLDVYVGGDYNARR
    ASLCEVLYSEAIYVMARPGHPLQAAALWEDVRYDWMVPPGSPTRSKLDNALIQ
    GGRQPAYRLESSMLANIELLRSDMLSTIGSARVVEHLAIGLVARLAEIPGAV
    GCMWRDEPHRASEDLIACREGAAAE"
    9181..10242
    /gene="gtda"
    /note="PA2470"
    9181..10242
    /gene="gtda"
    /codon_start=1
    /transl_table=11
    /product="gentisate 1,2-dioxygenase"
    /protein_id="AAG05858.1"
    /db_xref="GI:9948520"
    /translation="MHNDNSTEKRADEFRRIRQOHLPLMEALSLVPAOPAGCQAA
    LMRYRELPLFLEAGDLISAEAVRVLVLENPLPOSSITPLSYGLQILGELA
    PSRRHQSALRFVEYGVGAYNSVDCETRMEDGFTITPSMTWHDGNRAAEQGEY
    VWLDGDIPLMRFEGATFENHDEPVQPLRGESDLSARGSNMLPLRHOPTTSTSL
    FSTPYARSREALERLARERPPDWEGKRLRVNATGGMAPTATCQLDLPACFLQD
    PARSTDVSYSVSVSGVAQIGRRFAFEAKDLFVPSMAELRLEAGATDCLVFSFD
    RPVOAALGILRESRETLAH"
    10255..10953
    /gene="PA2471"
    10255..10953
    /gene="PA2471"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAG05859.1"
    /db_xref="GI:9948521"
    /translation="MNHLPFAEAPVSLTIHSGDQTFPVRRYCYGRRYAAHAREMGFD
    PERBPPEFCFPAADVPAAGSTLELAPSGTGNVRYEIELVAIAIGGCDPLLEDA
    EEHWGAYAGVGDMDTRDLOMRMRMGKPMETGKAFDRSADPIGPLYPASOVGHPHAAI
    SLDVGEDDRSDIDLOLMSVAETVSYLSRFELRPGDLVFTGPEGVGAERGERML
    GAIDGLSELISRVRY"
    BASE COUNT      1675 a      3678 c      3860 g      1808 t
    ORIGIN

Query Match      84.0%; Score 16.8; DB 1; Length 11021;
Best Local Similarity 90.0%; Pred. No. 3.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 99gcctgctgaacagggctc 20
      ||| || ||||| |||
Db      8834 GGGCGTGTGAACAGGCTC 8853

RESULT 6
AC013777/c      AC013777      156640 bp      DNA      HTG      12-MAR-2000
LOCUS          AC013777
DEFINITION     Homo sapiens clone Rpl1-12D3, WORKING DRAFT SEQUENCE, 3 unordered
               pieces
ACCESSION     AC013777
VERSION       AC013777.3      GI:7107917
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ADPHORS 1 (bases 1 to 156640)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome, clone RP11-1203
REFERENCE Unpublished
2 (bases 1 to 156640)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collimore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehocsky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL On Feb 28, 2000 this sequence version replaced gi:6539349.
COMMENT All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L3255
Center clone name: 12.D.3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147337 bases at least Q40
Consensus quality: 153442 bases at least Q30
Consensus quality: 155614 bases at least Q20
Insert size: 15500; agarose-fp
Insert size: 156440; sum-of-coverage
Quality coverage: 5.8 in Q20 bases; agarose-fp
Quality coverage: 5.8 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8361: contig of 8361 bp in length
* 8362 8461: gap of 100 bp
* 8462 54664: contig of 46203 bp in length
* 54665 54764: gap of 100 bp
* 54765 156640: contig of 101876 bp in length.
Location/Qualifiers
1. 156640
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1203"
/clone_lib="RPC1-11 Human Male BAC"
1. 8361
misc_feature
/note="assembly_fragment"
8462. 54664
misc_feature
/note="assembly_fragment"

misc_feature
54765..156640
/note="assembly_fragment"
clone_end=17
vector_side="left"
BASE COUNT 50566 a 27994 c 28634 g 49246 t 200 others
ORIGIN
Query Match 84.0%; Score 16.8; DB 2; Length 156640;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ggagctgctgacagcggc 20
Db 23691 GGCCTTGATGACAGGCGTC 23672
RESULT 7
AC027466
LOCUS Homo sapiens clone RP11-10H5, WORKING DRAFT SEQUENCE, 28 unordered
DEFINITION pieces.
ACCESSION AC027466
VERSION AC027466.2 GI:7677892
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ADPHORS 1 (bases 1 to 182110)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-10H5
REFERENCE Unpublished
2 (bases 1 to 182110)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,M., Baldwin,J., Barna,N., Bastien,V., Bida,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 2, 2000 this sequence version replaced gi:7342210.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L8378
Center clone name: 10_H_5
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168628 bases at least Q40
Consensus quality: 175797 bases at least Q30
Consensus quality: 178246 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 179410; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1059: contig of 1059 bp in length
* 1060 1159: gap of 100 bp
* 1160 2631: contig of 1472 bp in length
* 2632 2731: gap of 100 bp
* 2732 4878: contig of 2147 bp in length
* 4879 4978: gap of 100 bp
* 4979 5480: contig of 502 bp in length
* 5481 5580: gap of 100 bp
* 5581 6997: contig of 1417 bp in length
* 6998 7097: gap of 100 bp
* 7098 9964: contig of 2867 bp in length
* 9965 10064: gap of 100 bp
* 10065 13117: contig of 3053 bp in length
* 13118 13217: gap of 100 bp
* 13218 16682: contig of 3465 bp in length
* 16683 16782: gap of 100 bp
* 16783 20772: contig of 3990 bp in length
* 20773 20872: gap of 100 bp
* 20873 25281: contig of 4409 bp in length
* 25282 25381: gap of 100 bp
* 25382 31358: contig of 5977 bp in length
* 31359 31458: gap of 100 bp
* 31459 36260: contig of 4802 bp in length
* 36261 36360: gap of 100 bp
* 36361 42830: contig of 6470 bp in length
* 42831 42930: gap of 100 bp
* 42931 48864: contig of 5934 bp in length
* 48865 48964: gap of 100 bp
* 48965 56430: contig of 7466 bp in length
* 56431 56530: gap of 100 bp
* 56531 62471: contig of 5941 bp in length
* 62472 62571: gap of 100 bp
* 62572 68817: contig of 6246 bp in length
* 68818 68917: gap of 100 bp
* 68918 75405: contig of 6488 bp in length
* 75406 75505: gap of 100 bp
* 75506 82266: contig of 6761 bp in length
* 82267 82366: gap of 100 bp
* 82367 89219: contig of 6853 bp in length
* 89220 89319: gap of 100 bp
* 89320 94967: contig of 5648 bp in length
* 94968 95067: gap of 100 bp
* 95068 102013: contig of 6946 bp in length
* 102014 102113: gap of 100 bp
* 102114 111619: contig of 9506 bp in length
* 111620 111719: gap of 100 bp
* 111720 121055: contig of 9336 bp in length
* 121056 121155: gap of 100 bp
* 121156 132970: contig of 11815 bp in length
* 132971 133070: gap of 100 bp
* 133071 146270: contig of 13200 bp in length
* 146271 146370: gap of 100 bp
* 146371 157654: contig of 11284 bp in length
* 157655 157754: gap of 100 bp

FEATURES * 157755 182110: contig of 24356 bp in length.
source Location/Qualifiers
1. 182110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-10H5"
/clone_11p="RP11-11 Human Male BAC"
1. 1059
/note="assembly-fragment"
misc_feature
/note="assembly-fragment"
1160. 2631
/note="assembly-fragment"
misc_feature
2732. 4878
/note="assembly-fragment"
misc_feature
4979. 5480
/note="assembly-fragment"
clone end:SP6
vector side:left"
5581. 6997
/note="assembly-fragment"
7098. 9964
/note="assembly-fragment"
10065. 13117
/note="assembly-fragment"
13218. 16682
/note="assembly-fragment"
16783. 20772
/note="assembly-fragment"
20873. 25281
/note="assembly-fragment"
25382. 31358
/note="assembly-fragment"
31459. 36260
/note="assembly-fragment"
36361. 42830
/note="assembly-fragment"
42931. 48864
/note="assembly-fragment"
48965. 56430
/note="assembly-fragment"
56531. 62471
/note="assembly-fragment"
62572. 68817
/note="assembly-fragment"
68918. 75405
/note="assembly-fragment"
75506. 82266
/note="assembly-fragment"
82367. 89219
/note="assembly-fragment"
89320. 94967
/note="assembly-fragment"
95068. 102013
/note="assembly-fragment"
102114. 111619
/note="assembly-fragment"
111720. 121055
/note="assembly-fragment"
121156. 132970
/note="assembly-fragment"
133071. 146270
/note="assembly-fragment"
146371. 157654
/note="assembly-fragment"
157755. 182110
/note="assembly-fragment"
BASE COUNT 59635 a 31686 c 31475 g 56612 t 2702 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 182110;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 1 99gcttgctgaacagggc 20
|| ||||| ||||| |||||
Db 100300 GGCCTTGCAATGAACAGGGTC 100319

RESULT 8
AC068806/c
LOCUS
DEFINITION Mus musculus chromosome 15 clone RP23-244G13, *** SEQUENCING IN
PROGRES ***; 65 unordered pieces.
AC068806 184043 bp DNA HTG 10-NOV-2000
AC068806
AC068806.8 GI:11128321
HTG: HTGS-PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 184043)
Mezker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okunou,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Bunay,C., Bunac,C., Burkett,C., Chacko,U., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosack,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louisedge,H., Lozada,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mel,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 184043)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2000 this sequence version replaced g1:9929620.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAEX
Center clone name: RP23-244G13
----- Summary Statistics
Sequencing vector: M13; I08821
Chemistry: Dye-Primer Bodipy: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 118977 bases at least Q40
Consensus quality: 148815 bases at least Q30
Consensus quality: 160843 bases at least Q20
Estimated insert size: 161140; sum-of-ctg-estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.1x in Q20 bases; sum-of-ctg-estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8429: contig of 8429 bp in length
* 8430 8529: gap of unknown length

8530 13768: contig of 5239 bp in length
13769 13868: gap of unknown length
13869 19652: contig of 5784 bp in length
19653 19752: gap of unknown length
19753 23628: contig of 3876 bp in length
23629 30765: gap of unknown length
30766 30865: gap of unknown length
30866 35440: contig of 4575 bp in length
35441 41474: gap of unknown length
41475 41574: gap of unknown length
41575 45129: contig of 3555 bp in length
45130 45329: gap of unknown length
45330 49005: contig of 3776 bp in length
49006 49105: gap of unknown length
49106 53421: contig of 4316 bp in length
53422 53521: gap of unknown length
53522 58826: contig of 5305 bp in length
58827 62453: gap of unknown length
62454 62553: gap of unknown length
62554 65563: contig of 3010 bp in length
65564 65663: gap of unknown length
65664 69006: contig of 3343 bp in length
69007 72380: gap of unknown length
72381 72880: contig of 3274 bp in length
72881 76210: gap of unknown length
76210 76310: contig of 3730 bp in length
76311 79104: gap of unknown length
79105 79204: contig of 2794 bp in length
79205 82313: gap of unknown length
82314 82413: contig of 3109 bp in length
82414 85674: gap of unknown length
85675 85774: contig of 3261 bp in length
85775 88527: gap of unknown length
88528 88627: contig of 2753 bp in length
88628 91695: gap of unknown length
91696 91795: gap of unknown length
91796 96306: contig of 4411 bp in length
96307 96306: gap of unknown length
96307 98887: contig of 2581 bp in length
98888 98987: gap of unknown length
98988 102102: gap of unknown length
102103 102202: contig of 3115 bp in length
102203 104750: gap of unknown length
104751 104850: contig of 2548 bp in length
104851 107655: gap of unknown length
107656 107755: contig of 2805 bp in length
107756 110636: gap of unknown length
110637 110736: contig of 2861 bp in length
110737 113677: gap of unknown length
113678 113777: contig of 2941 bp in length
113778 116806: gap of unknown length
116807 116906: contig of 3029 bp in length
116907 119566: gap of unknown length
119567 119566: contig of 2660 bp in length
119567 119666: gap of unknown length
119667 121611: contig of 1945 bp in length
121612 121711: gap of unknown length
121712 124049: contig of 2338 bp in length
124050 124149: gap of unknown length
124150 126370: contig of 2221 bp in length
126371 126470: gap of unknown length
126471 128520: contig of 2050 bp in length
128521 128620: gap of unknown length
128622 131123: contig of 2503 bp in length
131124 131223: gap of unknown length
131224 134398: contig of 3175 bp in length
134399 134498: gap of unknown length
134499 137701: contig of 3203 bp in length
137702 137801: gap of unknown length
137802 139848: contig of 2047 bp in length

```
* 139849 139948: gap of unknown length
* 139949 141518: contig of 1570 bp in length
* 141519 141618: gap of unknown length
* 141619 143162: contig of 1544 bp in length
* 143163 143262: gap of unknown length
* 143263 145434: contig of 2172 bp in length
* 145435 145534: gap of unknown length
* 145535 148086: contig of 2552 bp in length
* 148087 148186: gap of unknown length
* 148187 149901: contig of 1715 bp in length
* 149902 150001: gap of unknown length
* 150002 151443: contig of 1442 bp in length
* 151444 151543: gap of unknown length
* 151544 153189: contig of 1646 bp in length
* 153190 153289: gap of unknown length
* 153290 155228: contig of 1939 bp in length
* 155229 155328: gap of unknown length
* 155329 157583: contig of 2255 bp in length
* 157584 157683: gap of unknown length
* 157684 159105: contig of 1422 bp in length
* 159106 159205: gap of unknown length
* 159206 160507: contig of 1302 bp in length
* 160508 160607: gap of unknown length
* 160608 162187: contig of 1580 bp in length
* 162188 162287: gap of unknown length
* 162288 164054: contig of 1767 bp in length
* 164055 164154: gap of unknown length
* 164155 165237: contig of 1083 bp in length
* 165238 165337: gap of unknown length
* 165338 167984: contig of 2647 bp in length
* 167985 168084: gap of unknown length
* 168085 169427: contig of 1343 bp in length
* 169428 169527: gap of unknown length
* 169528 170940: contig of 1413 bp in length
* 170941 171040: gap of unknown length
* 171041 173028: contig of 1988 bp in length
* 173029 173128: gap of unknown length
* 173129 174177: contig of 1049 bp in length
* 174178 174277: gap of unknown length
* 174278 175280: contig of 1003 bp in length
* 175281 175380: gap of unknown length
* 175381 176950: contig of 1570 bp in length
* 176951 177050: gap of unknown length
* 177051 178141: contig of 1091 bp in length
* 178142 178241: gap of unknown length
* 178242 179275: contig of 1034 bp in length
* 179276 179375: gap of unknown length
* 179376 180483: contig of 1108 bp in length
* 180484 180583: gap of unknown length
* 180584 181600: contig of 1017 bp in length
* 181601 181700: gap of unknown length
* 181701 182848: contig of 1148 bp in length
* 182849 182948: gap of unknown length
* 182949 184043: contig of 1095 bp in length.
```

```
FEATURES
    source
        1. 184043
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="15"
            /clone="RP23-244C13"
```

```
BASE COUNT  47869 a 40348 c 41235 g 47872 t 6499 others
```

```
Query Match      84.0%; Score 16.8; DB 2; Length 184043;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      1 ggctctgctgaacagggtc 20
        ||||| ||||| |||||
Db 110169 GGCTTGCTGAGCAGGTC 110150
```

```
RESULT  9
AC063109/c
```

```
LOCUS      AC063109      961 bp      DNA
DEFINITION  Giardia intestinalis clone NJ3102 strain WB-C6, LOW-PASS SEQUENCE
ACCESSION  AC063109
VERSION    AC063109.1 GI:7633768
KEYWORDS   HTG; HTGS_PHASEO.
SOURCE     Giardia intestinalis.
ORGANISM   Giardia intestinalis
REFERENCE  1 (bases 1 to 961)
AUTHORS    Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Giardia: a model for ancient eukaryotic genome analysis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 961)
AUTHORS    Nixon,J., Morrison,H.G., McArthur,A.G., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Direct Submission
JOURNAL    Submitted (21-APR-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT    * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
```

```
FEATURES
    source
        1. 961
            Location/Qualifiers
```

```
1. 961
    /organism="Giardia intestinalis"
    /strain="WB-C6"
    /db_xref="taxon:5741"
    /clone="NJ3102"
BASE COUNT  273 a 236 c 231 g 221 t
ORIGIN
```

```
Query Match      82.0%; Score 16.4; DB 2; Length 961;
Best Local Similarity 94.4%; Pred. No. 8.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 ggctctgctgaacagggtc 18
        ||||| ||||| |||||
Db      376 GAGCTTGCTGACAGCG 359
```

```
RESULT  10
```

```
AC052567/c  968 bp      DNA      HTG      15-APR-2000
LOCUS      Giardia intestinalis clone KJ3702 strain WB-C6, LOW-PASS SEQUENCE
DEFINITION  SAMPLING.
```

```
ACCESSION  AC052567
VERSION    AC052567.1 GI:7575887
KEYWORDS   HTG; HTGS_PHASEO.
SOURCE     Giardia intestinalis.
ORGANISM   Giardia intestinalis
REFERENCE  1 (bases 1 to 968)
AUTHORS    Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Giardia: a model for ancient eukaryotic genome analysis
JOURNAL    Unpublished
```

```
REFERENCE  2 (bases 1 to 968)
AUTHORS    Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Direct Submission
JOURNAL    Submitted (15-APR-2000) Josephine Bay Paul Center for Comparative
```


FEATURES

source

University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

Location/Qualifiers

1. 12357

/organism="Pseudomonas aeruginosa"

/strain="PA01"

/db_xref="taxon:287"

/complement(181..513)

/gene="PA0329"

/complement(181..513)

/gene="PA0329"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAG03718.1"

/db_xref="GI:9946177"

/translation="MAKFKLKKAKDGFHNLHAANGELIITLSELYKAKDSALGITE
SVKRSQRDGAFEVKRPNNGKFHFVLRKATNGQYVGSOLYASQANAEVGSVKRATP
EAGLSDES"

/complement(637..1308)

/gene="rpiA"

/note="PA0330"

/complement(637..1308)

/gene="rpiA"

/codon_start=1

/transl_table=11

/product="ribose 5-phosphate isomerase"

/protein_id="AAG03719.1"

/db_xref="GI:9946178"

/translation="MNODLKQAVAAVVDHILPHDSKIVGCTGTANFTDALA
RHKAFDGAVASSEATKRLKEHGIPIVELTVSELEFYVDGADESNNRLKGGGA
ALTRKIVAAVAKFTICTADASKLVPILGFPLEVPIMARSHVARKQLVKGDPVY
REGVLTDMGNTILDVHNLRIDSPVELEKINAIYGVVTNGLEAARPADLLILGTADV
KTLKA"

/complement(1566..3080)

/gene="l1va1"

/note="PA0331"

/complement(1566..3080)

/gene="l1va1"

/codon_start=1

/transl_table=11

/product="threonine dehydratase, biosynthetic"

/protein_id="AAG03720.1"

/db_xref="GI:9946179"

/translation="MLEQYVKRLITSRYVDVAVENTPLOPAROLSERLGNVLLKREDL
QPVSPFKRGAVNKVAQLTEEEKARGVIAASAGNHAGLALAKROGIRAVIYMPKT
PEIKQAVARHAGKAVILHGDAPEALAHALKLVDEKGYTFVHPDDPTIAGOGVAM
ELIKOQGRDLAIFVPGGGGIVAGIAAYVYLRBEIKVIGVPEDESCLQANAAAG
RVVLGQVGLPADYAVNAOIGQHTPDICDHDEVITVSTDEICAIKIDTDSITE
PAGALVAAGIKKIVEREKAEQOTLVAIDSGANVNDRLRHAERAELEGERREALIAY
IPEPRGSKAECAVNGKROITEFNRYHSGSEAHIFVGQTHPENDDPEALVAALREK
GFVPLDLDNELAKLIRHMVGHAVKSDSEMDVRFEPERPGALFNFLTLGGRWNI
SMFYRNHGADGDNVAVGLQVPEDEBRLIPQTLAIGPYMDETANAPYQLFLC"

/complement(3200..3667)

/gene="PA0332"

/note="PA0332"

/complement(3200..3667)

/gene="PA0332"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG03721.1"

/db_xref="GI:9946180"

/translation="MEHFLQIKILHGVATVLLFGLGLGLAFVANSRMTGDAARVARG
FKRVLIGMPLGLSLALPYSGMWLVHLGMPGOSMLLEGACTLYLIGCFMVLVLAG
RLAVOHGAGAEFAGAALEGBQTLREIKVGLVYALLGAASIAVAIVLMSKPL"

/complement(3718..4989)

/gene="PA0333"

/note="PA0333"

/complement(3718..4989)

/gene="PA0333"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG03722.1"

/db_xref="GI:9946181"

/translation="WRILVAGGFGVRHLLPALLAACHELLTARPPVAPAGVNR
LALDELRLAERPDSFAMPAGVDLLINAGTMSLDEASARVODSGARTLFDLAAHGA
KVLQISALGAGNHDPVAFPLASKAADRHLELIGIPALVLRSLIPGSGASSANLER
SPLIPLILDRARLOPLHVEDIVGAVALLRCQPERVORPLVGPOLNQCCELLDEL
RRAQGRPRGRIVVPPAALLDLGGIGRAGRRTTSPSLKLYVRHNDLADPLDBACG
YRCAPLASRLGWPQAAARSIALAKRPLMLAALVLIWGTIVACGPGGWLRLGEA
GHGMPASLAVIAGALDGLGVLGLLRMRRRALLAOFWMLMLGYSLAISILPHYV
DPMYAKNIVLWVATVLLDGEDETRBARG"

/complement(5426..6664)

/gene="PA0334"

/note="PA0334"

/complement(5426..6664)

/gene="PA0334"

/codon_start=1

/transl_table=11

/product="probable MFS transporter"

/protein_id="AAG03723.1"

/db_xref="GI:9946182"

/translation="MPTTPAPGSLRPVVIVLVTALCLAGDSMLYIALPIYHVEGLE
ALMQVGLSLINRLRLEPNLIGRLVKNHLAKTGLLAVVIGVLSAGVGLASGFLA
WLLRGLMIGMSFERRIGLSAVVYCAADHQRHAGMLYNGLYRLSGVGLGLGLV
FVLGIPALALSFALLALIGLPLAKGFDLPENPGHARTSAATSTVARRPPRGVY
ILSGPATALLIOGVLAATLSALIRSGTEVSLIGLALSAGLSGIIOALRMSWPMIL
AGRCGANSRDPGRGRRTVYAAVLIVGAMFGLSRPLMLALIALVLMVATRLTL
TDALADVARAGADVGEFMTRTYIVODIGALGLPALFLLEPGCFAMLVGSLAFL
VLALMLRRADPQGLTLTR"

/complement(6714..7367)

/gene="PA0335"

/note="PA0335"

/complement(6714..7367)

/gene="PA0335"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/db_xref="GI:9946181"

/translation="WRILVAGGFGVRHLLPALLAACHELLTARPPVAPAGVNR
LALDELRLAERPDSFAMPAGVDLLINAGTMSLDEASARVODSGARTLFDLAAHGA
KVLQISALGAGNHDPVAFPLASKAADRHLELIGIPALVLRSLIPGSGASSANLER
SPLIPLILDRARLOPLHVEDIVGAVALLRCQPERVORPLVGPOLNQCCELLDEL
RRAQGRPRGRIVVPPAALLDLGGIGRAGRRTTSPSLKLYVRHNDLADPLDBACG
YRCAPLASRLGWPQAAARSIALAKRPLMLAALVLIWGTIVACGPGGWLRLGEA
GHGMPASLAVIAGALDGLGVLGLLRMRRRALLAOFWMLMLGYSLAISILPHYV
DPMYAKNIVLWVATVLLDGEDETRBARG"

/complement(5426..6664)

/gene="PA0334"

/note="PA0334"

/complement(5426..6664)

/gene="PA0334"

/codon_start=1

/transl_table=11

/product="probable MFS transporter"

/protein_id="AAG03723.1"

/db_xref="GI:9946182"

/translation="MPTTPAPGSLRPVVIVLVTALCLAGDSMLYIALPIYHVEGLE
ALMQVGLSLINRLRLEPNLIGRLVKNHLAKTGLLAVVIGVLSAGVGLASGFLA
WLLRGLMIGMSFERRIGLSAVVYCAADHQRHAGMLYNGLYRLSGVGLGLGLV
FVLGIPALALSFALLALIGLPLAKGFDLPENPGHARTSAATSTVARRPPRGVY
ILSGPATALLIOGVLAATLSALIRSGTEVSLIGLALSAGLSGIIOALRMSWPMIL
AGRCGANSRDPGRGRRTVYAAVLIVGAMFGLSRPLMLALIALVLMVATRLTL
TDALADVARAGADVGEFMTRTYIVODIGALGLPALFLLEPGCFAMLVGSLAFL
VLALMLRRADPQGLTLTR"

/complement(6714..7367)

/gene="PA0335"

/note="PA0335"

/complement(6714..7367)

/gene="PA0335"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG03724.1"

/db_xref="GI:9946183"

/translation="MRALPDLMDTLAAGSDHDSWGERLQCRGLVDAEYQARDAFY
ADIVAGKLVDLAYOFTOAILGRTEMAOLETWHQFQOEVEITVLAKGELHREHA
AGDRLVITATNRFTVGTPIARLGEVETLATECWRMGRTGTGFDPCCQGGKVVRL
QRMULDENGLDLEGASFYSDSINDPLLEKVSRPVAVDPDRRLRAEARKGMPILSR"

/complement(7571..8050)

/gene="PA0336"

/note="PA0336"

/complement(7571..8050)

/gene="PA0336"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAG03725.1"

/db_xref="GI:9946184"

/translation="MIDSDGRPNVGIILANEAGQVLMARRINQANQFPOGGINDRE
TPEBALRELNEEVLGAGDVRLIACIRGMRLRPLPQRLVTRTHSQPLCIGOKQMFLL
RLMSDEARVRMDITSKEPFQGMWVSVMYPLGQVYTERREYRRALKELAPRLIARD"

/complement(8073..10352)

/gene="p1sb"

/note="PA0337"

/complement(8073..10352)

/gene="p1sb"

/codon_start=1

/transl_table=11

/product="phosphoenolpyruvate-protein phosphotransferase
p1sb"

/protein_id="AAG03726.1"

/db_xref="GI:9946185"

/translation="MLNTLRKRIQVQVNSAKDLKALGIIIVQRYEAMGTQVCSYLLD
TEORFVLMTEGLNKRISIGKVGVLGVTGREEPILNENAAHRYHAFAT
GEERYASFLLAPITIHHRWVGVLYVQOKERQORDEGEARLVTMSQALAVYHAEXT
GSTINGLCKGIDQAKFVCPGAPGVGKAAVVLPADLEVVPDQVDDIDAEIL
FKQALEGVRADMRALSSKLSQLEKERRALFDVLLMLDASIGNEVKRIIRQGMQO
GALROVYMEHVORFELMDDAVLRERASDVLDIGRLIAYIQEERKQMLYPEQTIYS
EELSPAMLGVEPGRGLVGLSVSGSHVAILARAGIPTVANGAVDLPYSKDGIDL
IYGSYHGEVYTNPSAELVROYSDVYAPESRELSGILARALRPLCTETIDGRMPLMYNG
LLADVARAQERGAEGVGLYRTVEVPMINDRFPSEKEDLATYRQDLSFHLPTVMRL
DIGGDKALSFPIREDNPLGMRGIRVTLQHPFLVQTRAMLKASGDLNLRLLPM
ISGTHLEDEALHLIRHAWGEVRDEGVDIAMPPIGMVEIRPAAYQTRIELAROVDFLSV

gene
CDS
10378..11508
/gene="PA0338"
10378..11508
/gene="PA0338"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="PA03727.1"
/db_xref="GI:9946186"
/translation="MRRLRPKRRRIDCLPFGHMPYERARDCRKMDEBARSKPPGLLP
LSREDLEKTLRYALEIYSDGIMDNINATNOVSRSAGWYLMGYPPHSLPESVETWKS
IHPEIDYPRWASRPAVYLDGSEPCYECRCRTYSGLYLMISDRGRFPVEPBERGPRRA
IGAHHEITHOKRLVLELQQRNEELPDNNRLJELVAETELHRYNOALASKMAEQR
LSEIDPLEYENRRKFEQCLHHEWRRQRHGRATVAVIDVDFKRINDLEFGHSTGR
LVVAFRLVASELREVDVLAHMGGEFTLLPETGLAALALAEERLRQVRVSSQSEMG
ERLTASFVGVLNDGETLIDLCLCRVDALYRAKQRDCVACC"
complement(11512..12267)
/gene="PA0339"
complement(11512..12267)
/gene="PA0339"
/codon_start=1

Query Match 82.0% Score 16.4 DB 1: Length 12357;
Best Local Similarity 94.4% Pred. No. 6.2e+02;
Matches 17: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 999ctgctgaacag9 18
|||||
Db 8739 GGGCGTGTGACAGCG 8722

RESULT 13
ATT1008/c
LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone T1008 (ESSA
project).
ACCESSION AL61746
VERSION AL61746.2 GI:7327833
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 83499)
Bevan, M., Pohl, T., Weizenecker, T., Bancroft, I., Mewes, H. W.,
Rudd, S., Lemcke, K. and Mayer, K. F. X.
Unpublished
2 (bases 1 to 83499)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (23-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7JY Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
On Mar 24, 2000 this sequence version replaced gi:7320707.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
Location/Qualifiers
1..83499
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
1..11148
misc_feature
EMBL:AL137189 for analysis and annotation"

gene
complement(join(9701..9775,9875..10169,10388..10506))
/gene="T1008_10"
9701..10506
/gene="T1008_10"
/number=1
complement(9701..9775)
/gene="T1008_10"
/number=1
complement(join(9701..9775,9875..10169,10388..10506))
/gene="T1008_10"
/note="similarity to several hypothetical proteins -
different bacteria"
/codon_start=1
/product="putative protein"
/protein_id="CAB81913.1"
/db_xref="GI:7320708"
/translation="MSSEELVLSPTTIDNDGKLEPRKTTMAGQYKKDISPLEYNYNP
EGCTVTLAVEDIDAPDPSGLVPTWTVVVDIPPEKNGLEEGSGNEDQTTGIREG
NDHKIPGWRGRLPESHGHRFQFKFLAALDDPKIGHITVTKERLLIAIEGHVYGEAIIITC
LA"
complement(9776..9874)
/number=1
complement(9875..10169)
/gene="T1008_10"
/number=2
complement(10170..10387)
/number=2
complement(10388..10506)
/gene="T1008_10"
/number=3
13362..17018
/gene="T1008_20"
join(13362..13863,14283..14560,14726..15597,15675..16272,
16358..16528,16701..17018)
/gene="T1008_20"
/note="first exon could be a separate protein
similarity to different hypothetical proteins - different
species
Contains ATP/GTP-binding site motif A (P-loop) AA234-241"
/codon_start=1
/product="putative protein"
/protein_id="CAB81914.1"
/db_xref="GI:7320709"
/translation="MDEFNLKRENPNSSTTSSSSSFHRRKSETGNTKRSTSTLS
TDPQSVARORRHRIIDRFKTLQSVGVGAAMPVSMLEAIVYKFLKAOIWHQNM
LLEFNHETTSCTCYSPGAGCGFKLIGYDDVAPINDYTSQGPLVADSKYTPWPG
SVIDEQRHVTYFKIRATRAHLRGHCNCLIGETEFADQREKMEVQIEESKNOTSPB
SIENDAKAKQIVLLIGPGSGKSTFCPTAMSSHRPWSRLQODIVNNGKAGTKAQCLK
MATDSLREGKSVFIDRCNLDEORSEFIKLGPEFYHAVALLEIPAOVCISRYKRTG
HEGNLOGGRAAAVYVNMKLOSKELPKVNGEFSRIMECYSDADVDAVNMVYKLGPMDTL
PSCGCGEKKLDTKSGPGIMKPFKVSALPASSSEANETTPRKADEMTANVAVSPVKIG
SADIVPTLAPFSITADPFDLEKASDIYKKAEEFLSKIGTARLVLYDLSRGSKIIS
LVKAKASQKNIDSAKFEFTFVGDITKLNSEGGLHCNVIANATNMWHLKPGGGVNAIIR
AAGPDLETATVRANVTLLPGKAVVYPLPTCPLNMSGITHHVHIVLPNNMNPBDML
ANNDTKCGCKTLREAVYSLFEGFLSVYVDQSLPKRSPQTAASDGEDIKEDSENNKY
KGSODKAVTNLSESESLIDRSGSKSGKSGMNTALALSHAMPREHRENVLELNN
IYVINDQYPRKARKHIVLIRQESIDGLEVPKREKLOLLOEHNNYGLKWPVRFQEDAS
LIFRLGTHVPSMQQLHVTISQFNSDSLKNKRHMSFTTSPFRSDVYIEVENSOG
KANVASDILLKGLRCNCRSAHNPRIKLSHVSCHSQSPDHLLONNRLVARET"
13362..13863
/gene="T1008_20"
/number=1
13864..14282
/gene="T1008_20"
/number=1
14283..14560
/gene="T1008_20"
/number=2
14561..14725
/gene="T1008_20"
/number=2
14726..15597
/gene="T1008_20"
/number=3

intron 15598..15674
/gene="T1008_20"
/number=3
15675..16272
/gene="T1008_20"
/number=4
16273..16357
/gene="T1008_20"
/number=4
16358..16528
/gene="T1008_20"
/number=5
16529..16700
/gene="T1008_20"
/number=5
16701..17018
/gene="T1008_20"
/number=6
complement(17542..17712)
/gene="T1008_30"
/number=1
17542..19683
/gene="T1008_30"
complement(join(17542..17712,17780..18086,18187..18837,
18915..19078,19165..19683))
/gene="T1008_30"
complement(join(17542..17712,17780..18086,18187..18837,
18915..19078,19165..19683))
/gene="T1008_30"
/note="strong similarity to pyruvate decarboxylase (EC
4.1.1.1) pcdl - Arabidopsis thaliana, PIR:T05315
Contains thiamine pyrophosphate enzymes signature
AA463-482"
/codon_start=1
/product="pyruvate decarboxylase-like protein"
/protein_id="CAB81915.1"
/db_xref="GI:7320710"
/translation="MDPKIGAIDTCKPTTGDISPSSNAVAITDASAPITTTSESTLG
RHLSTRVLQAGVTDFVSPGDFNLTLDLHIAPELNNIGCCNELNAGYADYARSR
GVGACVYTFYVGGSVLNAIAGAYSENLPVLCIYGPNSNDFGTRILHHTIGLPPDS
OELRCFOTVTCYQAVVNNLDEAHQIDKATATLAKSKPVYISCNLAAPHPFPA
DPVFDLTPRMSNTMGLEAAVEATLEFLNKAVKVMYGGPKLRARAKSEAFLEADS
GYPLAVPSTKGLVPEHNPHETIGTWGAVSTPFCSEIVESADAYIFAGPIENDSVG
VSLLKKEKALIVHEDRVVYVANGPTFGCVASDFRELAKEKVKENETAYENERIYV
EGKPLKCPGPELVNMAFOHIQKMLSEFVIAETGDSMFCQKLRKPCGGEFQM
OYSGISGVATLIGYAOATPEKRYLSFIIGSGSPVATODISTMIRNGOKAITEILNNG
GYTVEVLIHGGSPVAVIKNMVYTGIVDAIHNGECKCMTKYRIEELVEAITYATTEKK
DSLCFIEVIYHKDPTSKELLEMGSRVSAANGRPNPQ"
complement(17713..17779)
/number=1
complement(17780..18086)
/gene="T1008_30"
/number=2
complement(18087..18186)
/number=2
complement(18187..18837)
/gene="T1008_30"
/number=3
complement(18838..18914)
/number=3
complement(18915..19078)
/gene="T1008_30"
/number=4
complement(19079..19164)
complement(19165..19683)
/gene="T1008_30"
/number=5
complement(join(20596..20766,20848..21154,21427..22077,
22151..22314,22380..22865))
/gene="T1008_40"
complement(20596..20766)
/gene="T1008_40"

gene /number=1
20596..22865
/gene="T1008_40"
complement(join(20596..20766,20848..21154,21427..22077,
22151..22314,22380..22865))
/gene="T1008_40"
/note="strong similarity to pyruvate decarboxylase (EC
4.1.1.1) pcdl - Arabidopsis thaliana, PIR:T05315
Contains thiamine pyrophosphate enzymes signature
AA452-471"
/codon_start=1
/product="pyruvate decarboxylase-like protein"
/protein_id="CAB81916.1"
/db_xref="GI:7320711"
/translation="MDVRSILPSNCGVATITDSPAATATILGSSAATLGRHLVQAG
VTDIFVPGDFNLTLDLINPELNNIGCCNELNAGYADYARSRGVGACVYTFV
GGLSVLNAINAGAYSENLPVLCIYGPNSNDFGTRILHHTIGLPPDFSGELRCFQTVTC

Query Match 82.0%; Score 16.4; DB 8; Length 83499;
Best local similarity 94.4%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 3 gcttgcttgacagagtc 20
||| ||||| ||||| |||
Db 7019 GCTTACTTGACAGCGTC 7002

RESULT 14
ATF4D11/c
LOCUS ATF4D11 101715 bp DNA PLN 27-AUG-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F4D11 (ESSAII
project).
ACCESSION AL022537
VERSION AL022537.1 GI:3063690
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 101715)
Beyan, M., Benes, V., Rechmann, S., Borkova, D., Ansgorge, W.,
Hohnleisel, J., Mewes, H.W., Mayer, K.F.X. and Schueller, C.
Unpublished
2 (bases 1 to 101715)
EU Arabidopsis sequencing project.
REFERENCE Direct Submission
AUTHORS Submitted (16-APR-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.beyan@bsrc.ac.uk

FEATURES
source location/Qualifiers
1..101715
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(2154..2813)
/gene="F4D11.10"
/number=1
complement(join(2154..2813,2909..3257,3625..3755,
3855..4496))
/gene="F4D11.10"
/note="similarity to multiple exostosin 2, Homo sapiens"
/codon_start=1
/product="putative protein"
/protein_id="CAI8582.1"
/db_xref="GI:3063691"
/db_xref="SPTREMBL:O65522"
/translation="MLNLMVREHLVYVYCHQONLVAKRSQAKMLRFAGLVLFITTT
FQVNRITRYGCDIFSLSSQDKFPSPIDVSTPEYTLSSGPERLSSSSRSREVDDEEST
GLKEDHVIIGFDKNDTVGGHDSFVEDVKDKETLIDLPLQTKSSSNESSEKIVEDADIAIE

NIRKMEILESNDSPVDNLSEVKKFMNVNSGVVSIEMMNLHOSRTSHVSLKVR
SSTIDHLLYARTQIENPPLIENDPLHTPLTYMNLSPKRSYELEMEKTLKYYVREGK
RPTLHKPKYKGTIVASEGFMKOLKSSRFVTWKDRKALFLYLPSSKMLETLVYVPGS
HSDKNLLOFLKNYIDMISSKSPFNKKTGGSDHFLVACHDMAPSETRQIMACITACLN
SDVSGFVFGKDALPETTLIVPRPLKALGKVFVSOQLIAFAGGHHGLRPLLO
NWGNRNPDMDKIFSEIPKSKKSYMEEKSKKCYCICPKGHEVNSPRVEALFEVCVP
VLIIDNFPFPEFVLENNESFAVFLERDIPDLKLNIVSITEERYREMOBRVMYOKHF
LMSKPERDFPFHILHSIMYRVFQI"
2154. .4496
/gene="F4D11.10"
/complement(join(2154. .2813,2909. .3257,3625. .3755,
3855. .4496))
/gene="F4D11.10"
complement(2814. .2908)
/number=1
complement(2909. .3257)
/gene="F4D11.10"
/number=2
complement(3258. .3624)
/number=2
complement(3625. .3755)
/gene="F4D11.10"
/number=3
complement(3756. .3854)
/number=3
complement(3855. .4496)
/gene="F4D11.10"
/number=4
complement(join(6155. .6547,7111. .7176,7846. .8006,
8082. .8136))
/gene="F4D11.20"
complement(6155. .6547)
/gene="F4D11.20"
/number=1
6155. .8136
/gene="F4D11.20"
complement(join(6155. .6547,7111. .7176,7846. .8006,
8082. .8136))
/gene="F4D11.20"
/note="similarity to predicted proteins, Arabidopsis
thaliana
Contains Prenyl group binding site (CAAX box), [CFLA]
contains EST gb:AA395096, T42550"
/codon_start=1
/product="putative protein"
/protein_id="CA18583.1"
/db_xref="GI:3063692"
/db_xref="SPTRMBL:O65523"
/translation="MTSLHKTLPFONPTTLDOLAMCHTGRDMSMKLENIIDEGPSTV
AREVRLALPETPEEMKELGRSWSISALILPHVGNKTSPTTPRTGEMKHLKYSMTI
GRMGRLKDKOEKKKQETRRNAEIIAASVAAVAAVATASNAIAAEDEVEST
TVAAAMASAAALINSHCIEIAGEIGAGYNOIATVASSATNARTGDIWALTASATGI
IYCPFA"
complement(6548. .7110)
/number=1
complement(7111. .7176)
/gene="F4D11.20"
/number=2
complement(7177. .7845)
/number=2
complement(7846. .8006)
/gene="F4D11.20"
/number=3
complement(8007. .8081)
/number=3
complement(8082. .8136)
/gene="F4D11.20"
/number=4
9271. .11609
/gene="F4D11.30"
complement(join(9271. .9522,9555. .9704,9801. .9917,
10008. .10115,10239. .10336,10486. .10655,10768. .10844,
11004. .11083,11294. .11609))

/gene="F4D11.30"
complement(9271. .9522)
/gene="F4D11.30"
/number=1
complement(join(9271. .9522,9555. .9704,9801. .9917,
10008. .10115,10239. .10336,10486. .10655,10768. .10844,
11004. .11083,11294. .11609))
/gene="F4D11.30"
/note="similarity to predicted protein, Synechocystis sp.,
FIR2:S74814"
/codon_start=1
/product="putative protein"
/protein_id="CA18584.1"
/db_xref="GI:3063693"
/db_xref="SPTRMBL:O65524"
/translation="MQFNIRKFEFLMLITYCLTTEKCRYHFDGTPRRKFEQMYRVS
PEKRESFCFMYSVENPAPRQSLPLEVALYPRFLFGVAQLIGADKYLCOYEDOSH
FWGDRHELVLGNTEFSAVPKAKPNKVEPSEFNRRVSGFOATPFMHGHCIDGRTD
YAEIVKSAKWEYSTRPVYGMGDVGAKOKSTAGWPAAPVPEPHMOICMAGGLSTGWE
WGERFERPADPSYSEKNMGGGFPKPKFMVOCNVEGATGEVALTLGAGLQOLPEL
TYENALVYCVHYDGMKMEFVPWNGVRRRMSPMGWYTTAENENHVVLELARTNAC
PLRAPTFEVLATACRDSCTSELKQIWERLIXDSKGLKVLTPNKAVKEYERLML
TMQVYILETKSSMAAVEIGGPPWGTWKDTSNTPELKQALQVPLDLSALGLVPEF
KPGIL"
complement(9523. .9554)
/number=1
complement(9555. .9704)
/gene="F4D11.30"
/number=2
complement(9705. .9800)
/number=2
complement(9801. .9917)
/gene="F4D11.30"
/number=3
complement(9918. .10007)
/number=3
complement(10008. .10115)
/gene="F4D11.30"
/number=4
complement(10116. .10238)
/number=4
complement(10239. .10336)
/gene="F4D11.30"
/number=5
complement(10337. .10485)
/number=5
complement(10486. .10655)
/gene="F4D11.30"
/number=6
complement(10656. .10767)
/number=6
complement(10768. .10844)
/gene="F4D11.30"
/number=7
complement(10845. .11003)
/number=7
complement(11004. .11083)
/gene="F4D11.30"
/number=8
complement(11084. .11293)
/number=8
complement(11294. .11609)
/gene="F4D11.30"
/number=9
12431. .12512
/note="trna"
13108. .13310
/note="EST Z33951 matches to coordinates 13310 to 13108"
complement(join(13229. .13902,14091. .14559,14675. .15093,
15334. .15553,15604. .15685,16119. .16330,16495. .16608,
16689. .16787,17385. .17445,17519. .17685))
/gene="F4D11.40"
complement(13229. .13902)

/gene="F4D11.40"
/number=1
13229. .17685
/gene="F4D11.40"
complement(join(13229. .13902,14091. .14559,14675. .15093,
15334. .15553,15604. .15685,16119. .16330,16495. .16608,
16689. .16787,17385. .17445,17519. .17685))
/gene="F4D11.40"
/note="similarity to tom-1b protein, Gallus gallus
contains EST gb:233951"

Query Match 82.0%; Score 16.4; DB 8; Length 101715;
Best Local Similarity 94.4%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 gcttgctgacaggtc 20
|||||
Db 46217 CCTTGCTTGAGAGGCTC 46200

RESULT 15
AC016662/c
LOCUS
DEFINITION
AC016662
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC016662 108387 bp DNA PLN 19-JAN-2001
Arabidopsis thaliana chromosome 1 BAC F2P9 genomic sequence,
complete sequence.
AC016662.7 GI:12325130
HTG.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 108387)
Lin,X., Kaul,S., Town,C.D., Benito,M., Cressy,T.H., Haas,B.J.,
Wu,D., Maitl,R., Konning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F2P9 genomic sequence
Unpublished
2 (bases 1 to 108387)
Lin,X. and Kaul,S.
Direct Submission
Submitted (04-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 108387)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280816.
Address all correspondence to:ateltigr.org

BAC clone F2P9 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GenScan⁺ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant
of GlimmerM, see Mihaela Perlea,
http://www.tigr.org/softlab/glimmer_hmm/glimmer.html, and
GeneSplicer (Mihaela Perlea and Steven Salzberg, contact
mperlea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tcdb/cgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).

Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
source
organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F2P9"
join(<35. .663,669. .>1255)
/gene="F2P9.1"
35. .1255
/gene="F2P9.1"
join(<35. .663,669. .1255)
/gene="F2P9.1"
/codon.start=2.
/product="unknown protein, 5' partial: 35-1255"
/protein_id="AAG52507.1"
/db_xref="GI:12325131"
/translation="KRILANLQRYGPPSSQHFSLRNHNLNSDLVETETSLGIL
LSLKIPKSLVSLSDCPNVLRSFELRKMRVPLNSCGHGVSSAIVSHSSRIG
IGPKPNCVAVPLKSLGFCOSTVSRILSSPPGVLLVNEIEIRKRIEFLVIGTARDNI
ERFPRVPEVYIGSTERPLKPLDDEPKMGSKDYKKEIKAREEPLVMSLPCDEL
INTLKCEVIRVSLTISGCAFAGFEVLRVDCICKYGLIRDAKVVWKPEPVLTLEI
EDIERKIEFLNRMGFHINCLADVPEYLGVLQIVPRYVNIYDILKLGGLGCDIGL
KGLIKPSMKRFYNLYVMPYPCERIFGKRKENVAVNRHHPAGLWKLKMPSNLTTKED
VONKSPFIESLA"
2203. .2222
/rpt_family="(A)n"
join(<4810. .5811,5900. .>6511)
/gene="F2P9.2"
4810. .6511
/gene="F2P9.2"
/note="similar to cytochrome P-450 GB:AAB37231 from
[Phlaeopsis sp. SM9108]"
join(4810. .5811,5900. .6511)
/gene="F2P9.2"
/codon.start=1
/product="putative cytochrome P-450; 4810-6511"
/protein_id="AAG52514.1"
/db_xref="GI:12325138"
/translation="MTIDMYLSFASRSGSSPPPSLELCISFLFISLFEMLPPGCA
WALKARFRHRESKGTGPAIPGSLPIFGLLAFVNNALFHLRIANLADCKAKALM
AFSGSTFRFVTSSEPTAKELINSASADRVKSAVELEFDRAMGAPGQDYWEIRL
RISSTHLESPKRISSGSESRKIGQNNVEIKNAKCEYGEVHIKILHFGSLNNVSS
VFGKTYNNEGIVYKSESNLEHLVSGEELGIFNNSDHPGGRMLDLOGVRKCS
LNGRVNVFVCKTIINDHKSRSIRDNPESEYDDQFVIVLGMGNSKLSQDMATVLM
EMTRGTDVAILEWILARVLPDIOAKAQADIDCTVGDGSGQVYDSLPKIPYRA
AIVKETLRMHPGFLLSMARLSDHTDIOGTHFTAGTTAANNMALTIDEVWPEAHE
YKPEFLGAGQESNNEPIMGSDRLAPGAGRCVPGKSMGLATVELWLAOLIGSYKW
SCGEVDSLSETLKLSLEMKNTLVCAIPRG"
complement(6762. .6783)
/rpt_family="AT-rich"
/db_xref="AT-rich"
/gene="F2P9.3"
6763. .8831
/gene="F2P9.3"
6763. .8831
/note="similar to flavonol 4'-sulfotransferase GB:P52837
from [Flaveria chloraefolia]"
7673. .8689
/gene="F2P9.3"
/codon.start=1
/product="putative flavonol sulfotransferase; 7673-8689"
/protein_id="AAG52512.1"
/db_xref="GI:12325136"
/translation="MESKTTQNSGEVELTEFEKTOKKYODFIATLPKSGWRDELL
TOYGGHMOQCELEGLPHARDHEAPATDVLVCSYPTGTTWALKALVAIVNRSYSD
AANPLLRNCHHEFVPEIDFAFVYQDRKNPLSTTHIPKALIPDSIVNSGCM
VYIMRDKDFEISWMTFLHKEKSGEGLASLEDSFDMCEGLISYGGVLDHVGYNKA
YQENPDRIILRYETMRANPLPKYRCLAEPMGCGTFDEDEENGAAKVVVLCSPETLK
MLEANKGDKEREDRPAYANSATFRKKCVDMANYLTLPMAARIDGLVEERKPDTCIL
QHDH"
complement(<9123. .>10175)

```
gene      /gene="F2P9.4"
          complement(9123..10175)
          /gene="F2P9.4"
          /note="similar to FLAVONOL 4'-SULFOTRANSFERASE GB:P52837
          from [Flaveria chloraefolia]"
          complement(9123..10175)
          /gene="F2P9.4"
          /codon_start=1
          /product="putative flavonol sulfotransferase; 10175-9123"
          /protein_id="ANG52515.1"
          /db_xref="GI:12325139"
          /translation="MESETLTAKATITTTTLPSHDETKTESTEPEKNOKRYODLISTF
          PHEKWRPEKPEPLIEYGGWMLPSLEGCIHAQEPFOARPSDFIVCSYPTGTTWLKAL
          TPAIANRSRFDSSNPDLKRNDFEFVYIEPPEEDVDLKDGNLTFFSHIPEYL
          LPDSYKSGCKKVVYIMREPKDFTISMTFLIKERTDIEGVSNSLSESPDMFCGLSGYG
          PYLNHIIIAVKRAYQENPDRIPLFKETPMRADLPYVKSLSAEFMGIGFTAEERKQVE
          KVNLCSEPTLKNLEANKGKREDRDPGVANSAYFRKGVDGWSNYLTPEMARIDG
          LMEERFKGTGLEHNGK"
          complement(12228..12256)
          /rpt_family="AT_rich"
          complement(12462..12498)
          /rpt_family="AT_rich"
          complement(12549..12688)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..13136)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..12856)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..13052)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..12660)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..12632)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12549..13108)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12551..13080)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12551..12716)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12551..12912)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12551..12736)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12556..13100)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12562..12940)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12573..12660)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12579..13126)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          complement(12590..12968)
          /rpt_family="mini satellite from T32N15 (6329 to 6981 23
          copies of 28-bp tandem repeat)"
          repeat_region
```

```
repeat_region complement(12601..12688)
               /rpt_family="mini satellite from T32N15 (6329 to 6981 23
               copies of 28-bp tandem repeat)"
Query Match      82.0%; Score 16.4; DB 8; Length 108387;
Best Local Similarity 94.4%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Search completed: December 8, 2001, 10:50:54
Job time: 4919 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:21:35 ; Search time 2889.21 Seconds
(without alignments)
74.386 Million cell updates/sec

Title: US-09-508-147-23

Perfect score: 20

Sequence: 1 ggcgttgctggcgttgctt 20

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthm:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_prod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	20	100.0	609	11	BF529304 602041771
2	18.4	92.0	907	13	BF529304 602041771
3	18.4	92.0	907	13	BF529304 602041771
4	18.4	92.0	907	13	BF529304 602041771
5	17.4	87.0	680	13	BF529304 602041771
6	17.4	87.0	680	13	BF529304 602041771
7	17.4	87.0	680	13	BF529304 602041771
8	17.4	87.0	680	13	BF529304 602041771
9	16.8	84.0	279	13	BF529304 602041771
10	16.8	84.0	279	13	BF529304 602041771
11	16.8	84.0	279	13	BF529304 602041771
12	16.8	84.0	279	13	BF529304 602041771

Result No.	Score	% Query Match	Length	DB ID	Description
13	16.8	84.0	415	11	BF392364 UT-R-CA0-
14	16.8	84.0	439	13	A2188645 SP_1012_B
15	16.8	84.0	443	10	BE292335
16	16.8	84.0	458	11	BF555593
17	16.8	84.0	467	11	BF260902
18	16.8	84.0	468	13	A2408608
19	16.8	84.0	496	13	A2741652
20	16.8	84.0	507	11	BF411579
21	16.8	84.0	586	10	BE158993
22	16.8	84.0	587	10	AA902009
23	16.8	84.0	631	13	A2966594
24	16.8	84.0	674	13	A2730549
25	16.8	84.0	678	13	A0640140
26	16.8	84.0	784	11	BC430552
27	16.8	84.0	784	11	BC430552
28	16.8	84.0	969	11	BF312299
29	16.8	84.0	1061	13	CNS06R2N
30	16.8	84.0	1135	11	BC398842
31	16.4	82.0	143	13	BC112072
32	16.4	82.0	327	11	AZ393249
33	16.4	82.0	360	11	BI200538
34	16.4	82.0	386	10	AA705846
35	16.4	82.0	391	11	BC895992
36	16.4	82.0	403	11	BC146140
37	16.4	82.0	428	11	TC9148
38	16.4	82.0	445	10	AA021160
39	16.4	82.0	446	13	A2738038
40	16.4	82.0	485	13	AQ280223
41	16.4	82.0	517	13	A2758178
42	16.4	82.0	539	13	A2322302
43	16.4	82.0	546	13	A2344654
44	16.4	82.0	559	11	BF079083
45	16.4	82.0	573	10	BE235160

ALIGNMENTS

RESULT 1
BF529304 609 bp mRNA
LOCUS 602041771 NC1 CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4179641
DEFINITION 5', mRNA sequence.
ACCESSION BF529304
VERSION BF529304.1 GI:11616667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9489 row: m column: 18
High quality sequence stop: 500.
location/Qualifiers
1. 609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4179641"
/clone_lib="NC1 CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q loss"

```
/lab_host="DH10B (T1 phage-resistant)"
/Note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI-CGAP library."

BASE COUNT      116 a      204 c      177 g      112 t
ORIGIN

Query Match      100.0%; Score 20; DB 11; Length 609;
Best Local Similarity 100.0%; Pred. NO. 89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggagctgctggagctgctt 20
|||||
Db 587 GGGCTTGCTGGGCTTGCTT 568

RESULT 2
CNS01JUS/c
LOCUS
DEFINITION
CNS01JUS 907 bp DNA GSS 12-JUN-2001
Anopheles gambiae GSS SP6 end of clone 11F08 of Notredame1 library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION
AL145894.1 GI:7004056
VERSION
AL145894
KEYWORDS
ORGANISM
Anopheles gambiae
African malaria mosquito.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
1 (bases 1 to 907)
Genoscope.
REFERENCE
AUTHORS
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 907)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
1..907
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="Notredame1"
/Note="end : SP6"
BASE COUNT      189 a      235 c      254 g      223 t      6 others
ORIGIN

Query Match      92.0%; Score 18.4; DB 13; Length 907;
Best Local Similarity 95.0%; Pred. NO. 4.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggagctgctggagctgctt 20
|||||
Db 171 GGGCTTGCTGGGCTTGCTT 152

RESULT 3
CNS01JUS/c
LOCUS
DEFINITION
CNS01JUS 927 bp DNA GSS 12-JUN-2001
Anopheles gambiae GSS T7 end of clone 14G02 of Notredame1 library
from strain PEST of Anopheles gambiae (African malaria mosquito),
```

```
genomic survey sequence.
ACCESSION
AL147502
AL147502.1 GI:7005648
VERSION
GSS.
KEYWORDS
ORGANISM
Anopheles malaria mosquito.
African malaria gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
1 (bases 1 to 927)
Genoscope.
REFERENCE
AUTHORS
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 927)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Direct Submission
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
1..927
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="14G02"
/clone_lib="Notredame1"
/Note="end : T7"
BASE COUNT      201 a      240 c      245 g      204 t      37 others
ORIGIN

Query Match      92.0%; Score 18.4; DB 13; Length 927;
Best Local Similarity 95.0%; Pred. NO. 4.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggagctgctggagctgctt 20
|||||
Db 209 GGGCTTGCTGGGCTTGCTT 190

RESULT 4
BF845439
LOCUS
DEFINITION
BF845439 551 bp mRNA EST 13-JAN-2001
ACCESSION
BF845439
BF845439.1 GI:1201940
VERSION
BF845439.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 551)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
```

Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?Cl=RC4&t2=RC4-HT1109-311200-022-g12&t3=2000-12-31&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 118.
Location/Qualifiers

FEATURES

source

1..551

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT1109"
/dev_stage="Adult"/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT

87 a 133 c 204 g 127 t

ORIGIN

Query Match 90.0%; Score 18; DB 11; Length 551;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ggctgtctgggctgtct 19
|||||

Db 124 GGGCTGCTGGGCTGCT 141

RESULT 5

A0626238

LOCUS A0626238 332 bp DNA GSS 16-JUN-1999
DEFINITION CITBI-E1-2538N21.TR CITBI-E1 Homo sapiens genomic clone 2538N21.
DNA sequence.

ACCESSION A0626238

VERSION A0626238.1 GI:5088630

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 332)

Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.

Use of BAC End Sequences from Caltech Libraries for Sequence-Ready

Map Building

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetlgr.org

Clones are available from Research Genetics (Inforesgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

FEATURES

source

1..332

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2538N21"

/clone_lib="CITBI-E1"

/sex="male"

/cell_type="sperm"

/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

Query Match 87.0%; Score 17.4; DB 13; Length 332;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ggctgtctgggctgtct 19
|||||

Db 149 GGGCTGCTGGGCTGCT 167

RESULT 6

A0357322

LOCUS A0357322 680 bp DNA GSS 24-JAN-1999
DEFINITION CITBI-E1-2538M22.TR CITBI-E1 Homo sapiens genomic clone 2538M22.
DNA sequence.

ACCESSION A0357322

VERSION A0357322.1 GI:4184495

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 680)

Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.

Use of BAC End Sequences from Caltech Libraries for Sequence-Ready

Map Building

Unpublished (1997)

Other_GSS: CITBI-E1-2538M22.TF

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetlgr.org

Clones are available from Research Genetics (Inforesgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

FEATURES

source

1..680

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2538M22"

/clone_lib="CITBI-E1"

/sex="male"

/cell_type="sperm"

/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 160 a 182 c 162 g 176 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 680;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ggctgtctgggctgtct 19
|||||

Db 149 GGGCTGCTGGGCTGCT 167

RESULT 7

CNS06065/c

LOCUS CNS06065 989 bp DNA GSS 06-JUL-2001
DEFINITION T7 end of clone AX0AA005D03 of library AX0AA from strain CBS 7064

of *Pichia farinosa*, genomic survey sequence.
 AL15443
 VERSION AL15443.1 GI:12194651
 KEYWORDS GSS.
 SOURCE *Pichia farinosa*.
 ORGANISM *Pichia farinosa*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; *Pichia*.
 REFERENCE 1 (bases 1 to 989)
 de Montigny, J., Spéner, C., Souciet, J., Tekala, F., Dujon, B.,
 Wincker, P., Artiguenave, F. and Potier, S.
 Genomic exploration of the hemiascomycetous yeasts: 15. *Pichia*
sorbitophila
 FEMS Lett. 487 (1), 87-90 (2000)
 JOURNAL MEDLINE 20584725
 REFERENCE 2 (bases 1 to 989)
 Souciet, J., Aigle, M., Artiguenave, F., Blandin, G.,
 Boitlot-Fukuhara, M., Bon, E., Bottier, P., Casaregola, S.,
 de Montigny, J., Dujon, B., Durrens, P., Leplingle, A., Llorente, B.,
 Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
 Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEMS Lett. 487 (1), 3-12 (2000)
 JOURNAL MEDLINE 20584711
 REFERENCE 3 (bases 1 to 989)
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source
 1..989
 /organism="Pichia farinosa"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0A005D03"
 /clone_lib="AX0AA"
 /note="end : 17"
 BASE COUNT 251 a 219 c 278 g 238 t 3 others
 ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 989;
 Best Local Similarity 94.7%; Pred. No. 1.1e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ggcctgctgggctgctt 20
 |||||||
 Db 514 ggccttgcctttggccttgcctt 496

RESULT 8
 B1410123/c 1040 bp mRNA EST 14-AUG-2001
 DEFINITION 602964059F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5119479 5',
 mRNA sequence.
 ACCESSION B1410123
 VERSION B1410123.1 GI:15171046
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1040)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1AM1291 row: e column: 16
 High quality sequence stop: 892.

FEATURES
 source
 1..1040
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5119479"
 /clone_lib="NCI_CGAP_Lu33"
 /tissue_type="pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker. Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - 01190(dT) primer [5'
 TGTTCACCAATCTGAAGTGGAGCGCGCCCTCTCTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 197 a 315 c 344 g 184 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 11; Length 1040;
 Best Local Similarity 94.7%; Pred. No. 1.1e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ggcctgctgggctgctt 20
 |||||||
 Db 945 ggccttgcctttggccttgcctt 927

RESULT 9
 A2077069/c 192 bp DNA GSS 31-MAR-2000
 LOCUS RPCI-23-399M22.TJ RPCI-23 Mus musculus genomic clone RPCI-23-399M22
 DEFINITION 'DNA sequence.
 A2077069
 ACCESSION A2077069.1 GI:7369968
 VERSION A2077069
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 192)
 Zhao, S., Niernann, W., Felblyum, T., Malek, J., Shatsman, S., Akaret
 , B., Levins, M., Mogan, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-399M22.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edjlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Reseach Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 399 row: M column: 22
Seq primer: SP6
Class: BAC ends.

Location/Qualifiers

1. 192
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-399M22"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1: EcORI; Site 2: EcORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcORI and EcORI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 74 a 63 c 55 g 0 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 192;
Best Local Similarity 90.0%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 99gctgctgggcttgcct 20
||||| ||||||| |||||||
Db 168 GGGCTTGGCTTGGCTTCTT 149

RESULT 10
A2624523 279 bp DNA GSS 13-DEC-2000
LOCUS A2624523
DEFINITION M0463G15F Mouse 10kb plasmid U06C1M library Mus musculus genomic
clone U06C1M0463G15 F, DNA sequence.
ACCESSION A2624523
VERSION A2624523.1 GI:11746713
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 279)
REFERENCE 1 (bases 1 to 279)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0463 row: G column: 15
Seq primer: CGTGTAAACGACGCCACGT
Class: plasmid ends

High quality sequence stop: 279.
Location/Qualifiers

1. 279
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0463G15"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (9114732114|9b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 61 a 44 c 61 g 113 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 279;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 99gctgctgggcttgcct 20
||||| ||||||| |||||||
Db 68 GGTCTTGGCTTGGCTTCTT 87

RESULT 11
A2822766 386 bp DNA GSS 20-FEB-2001
LOCUS A2822766
DEFINITION 2M0096118F Mouse 10kb plasmid U06C1M library Mus musculus genomic
clone U06C2M0096118 F, DNA sequence.
ACCESSION A2822766
VERSION A2822766.1 GI:12992674
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 386)
REFERENCE 1 (bases 1 to 386)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0096 row: I column: 18
Seq primer: CGTGTAAACGACGCCACGT

Class: plasmid ends
High quality sequence stop: 386.
Location/Qualifiers

```
1. 386
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0096118"
/clone_11b="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g114732114(gd)AF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

BASE COUNT 103 a 90 c 87 g 106 t
ORIGIN

Query Match 84.0% Score 16.8; DB 13; Length 386;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0y 1 gggcttgctggctgctt 20
||||| ||| ||| ||| |||
Db 42 GGGCTTGCTGAGCTGCTT 61

RESULT 12
AM528369
LOCUS 391 bp mRNA EST 06-MAR-2000
DEFINITION UI-R-BT1-ajx-f-08-0-UI-s1 UI-R-BT1 Rattus norvegicus cDNA clone
ACCESSION AM528369
VERSION AM528369.1 GI:7170783
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msquares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the
normalized corpus-striatum library cDNA library preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 6-108, >BI-F*51NE/Alu
121-157, >AT-rich#low_complexity 323-366, >BI_MM51NE/Alu
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source location/Qualifiers

```
1. 391
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT1-ajx-f-08-0-UI"
/clone_11b="UI-R-BT1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; The library
UI-R-BT1 is a subtracted library derived from a mixture of
the following tissues: hippocampus, thalamus, mid-brain,
medulla, corpus striatum, cerebral cortex and testis. For
a detailed description of the library from which this
clone was derived, please visit our web site at
rategen.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG LIB=UI-R-BT1
TAG TISSUE=corpus-striatum
TAG_SEQ=CTAGC"
```

BASE COUNT 85 a 85 c 82 g 139 t
ORIGIN

Query Match 84.0% Score 16.8; DB 10; Length 391;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0y 1 gggcttgctggctgctt 20
||||| ||| ||| ||| |||
Db 214 GTGCTTGCTTGCTGCTT 233

RESULT 13
BF392364
LOCUS 415 bp mRNA EST 27-NOV-2000
DEFINITION UI-R-CAO-bf1-c-07-0-UI-s1 UI-R-CAO Rattus norvegicus cDNA clone
ACCESSION BF392364
VERSION BF392364.1 GI:11377221
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 415)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msquares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the normalized hippocampus library. CDNA Library Preparation: M.B. Soares lab clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-28, >AT-rich#Low-complexity 194-279, >BL-MACHINE/Alu Seq primer: M13 Forward POLYA=yes.

FEATURES

SOURCE

Location/Qualifiers
1. 415
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-bfi-c-07-0-UI"
/clone_lib="ui-r-ca0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0 library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_Lib=UI-R-CA0
TAG_Tissue=hippocampus
TAG_SEQ=GATTG

BASE COUNT 88 a 107 c 82 g 138 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 415;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcttgcctggcttgcctt 20
1 ||||||| |||||||
Db 85 GGCCTTGCTTGCTTGCTT 104

RESULT 14
A2188645/c 439 bp DNA GSS 30-AUG-2000
LOCUS
DEFINITION SP_1012_B1_E05.SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1012 Col=9 Row=J, DNA sequence.
ACCESSION A2188645
VERSION
KEYWORDS
SOURCE GSS
ORGANISM Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 439)
Cameron, R.A., Mahaitas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, R., Britten, R.J., Davidson, E.H. and Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1012 row: J column: 9

Seq primer: SP6
Class: BAC ends
High quality sequence stop: 439.
Location/Qualifiers
1. 439

FEATURES
SOURCE
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=1012 Col=9 Row=J"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACs3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 126 a 131 c 83 g 99 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 439;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcttgcctggcttgcctt 20
1 ||||||| |||||||
Db 108 GGCCTTGATTGGCCTTGATT 89

RESULT 15
BE292335 443 bp mRNA EST 13-JUL-2000
LOCUS
DEFINITION 601084092P1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3498287 5', mRNA sequence.
ACCESSION BE292335
VERSION BE292335.1 GI:9174656
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 443)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: <http://image.llnl.gov>
Plate: LAM8553 row: g column: 24
High quality sequence start: 22
High quality sequence stop: 379.
Location/Qualifiers
1. 443

FEATURES
SOURCE
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3498287"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue-type="infiltrating ductal carcinoma"
/dev-stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 75 a 128 c 114 g 126 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 443;

Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 99gcttgctt99gcttgctt 20
||||| ||||| |||||
Db 202 GGCTTGGTTGGCTTGGCTT 221

Search completed: December 8, 2001, 10:21:38
Job time: 3178 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:53:14 : Search time 136.23 Seconds
(without alignments)
33.249 Million cell updates/sec

Title: US-09-508-147-23

Perfect score: 20

Sequence: 1 ggccttgctggccttgctt 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*

6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.4	82.0	2424	1	US-08-821-119-16
C 2	16	80.0	1604	4	US-09-260-843-1
C 3	15.8	79.0	9323	1	US-08-038-682-6
C 4	15.8	79.0	9323	1	US-08-302-832-6
C 5	15.8	79.0	9323	2	US-08-530-198-6
C 6	15.8	79.0	9323	2	US-08-469-880-6
C 7	15.8	79.0	9323	2	US-08-728-470-6
C 8	15.8	79.0	9323	2	US-08-617-697-6
C 9	15.8	79.0	9323	4	US-08-719-641-6
C 10	15.4	77.0	1405	1	US-08-390-162-3
C 11	15.4	77.0	1405	1	US-08-685-945B-3
C 12	15.4	77.0	2107	1	US-08-390-162-1
C 13	15.4	77.0	2107	1	US-08-685-945B-1
C 14	14.8	74.0	271	1	US-08-306-691B-24
C 15	14.8	74.0	271	5	PCT-US93-06251-66
C 16	14.8	74.0	1611	1	US-07-820-011A-3
C 17	14.8	74.0	1611	5	PCT-US93-00445-3
C 18	14.8	74.0	12687	1	US-08-676-169-1
C 19	14.8	74.0	12687	1	US-08-981-459-1
C 20	14.4	72.0	20	2	US-08-910-629A-41
C 21	14.4	72.0	20	3	US-09-287-796-41
C 22	14.4	72.0	20	4	US-09-130-616-41
C 23	14.4	72.0	1780	5	PCT-US94-12913A-17
C 24	14.4	72.0	1782	3	US-09-209-668-16
C 25	14.4	72.0	1782	4	US-09-130-616-169
C 26	14.4	72.0	2621	2	US-08-553-619B-8
C 27	14.4	72.0	51259	3	US-08-781-891-209

C 28	14.2	71.0	851	4	US-09-008-892-10	Sequence 10, Appl
C 29	14.2	71.0	1128	2	US-08-459-101A-1	Sequence 1, Appl
C 30	14.2	71.0	2079	4	US-09-381-849-4	Sequence 4, Appl
C 31	14.2	71.0	3461	2	US-08-389-564B-1	Sequence 1, Appl
C 32	14.2	71.0	3461	3	US-08-466-047B-1	Sequence 1, Appl
C 33	14.2	71.0	4081	4	US-08-999-774A-1	Sequence 1, Appl
C 34	14.2	71.0	4164	1	US-08-188-582-3	Sequence 3, Appl
C 35	14.2	71.0	4164	1	US-08-646-715-3	Sequence 3, Appl
C 36	14.2	71.0	5102	1	US-08-494-168-1	Sequence 1, Appl
C 37	14.2	71.0	1843	4	US-09-078-294-6	Sequence 6, Appl
C 38	14.2	71.0	90050	4	US-09-245-041-5	Sequence 5, Appl
C 39	14.2	71.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 40	14	70.0	768	4	US-09-328-111-667	Sequence 667, App
C 41	13.8	69.0	332	2	US-08-619-362A-15	Sequence 1, Appl
C 42	13.8	69.0	388	1	US-08-121-063-1	Sequence 1, Appl
C 43	13.8	69.0	477	2	US-08-951-648-38	Sequence 38, Appl
C 44	13.8	69.0	477	3	US-09-174-437-38	Sequence 38, Appl
C 45	13.8	69.0	1347	4	US-09-412-102-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-821-119-16/c
Sequence 16, Application US/08821119
Patent No. 5821104
GENERAL INFORMATION:
APPLICANT: Holm, Kaj Andre
APPLICANT: Rasmussen, Grethe
APPLICANT: Halkier, Torben
APPLICANT: Lehmedek, Tan
TITLE OF INVENTION: [tripeptidyl Amino peptidase]
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: No. 58211040 No. 5821104disk of No. 5821104th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,119
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2424 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
US-08-821-119-16
Query Match 82.0% Score 16.4; DB 1; Length 2424;
Best Local Similarity 94.4%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
2 ggccttgctggccttgctt 19
|||||
```

Db 815 GGCGTGTGGGCTTGCT 798

RESULT 2

US-09-260-843-1/c

Sequence 1, Application US/09260843

Patent No. 6271439

GENERAL INFORMATION:

APPLICANT: Johal, Gurmukh S

APPLICANT: Briggs, Steven P

APPLICANT: Gray, John

APPLICANT: Hu, Gongshe

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL DEATH AND

TITLE OF INVENTION: ENHANCING DISEASE RESISTANCE TO PLANT PATHOGENS

FILE REFERENCE: Pioneer 035718/175368

CURRENT FILING DATE: 1999-03-02

EARLIER APPLICATION NUMBER: 60/076754

EARLIER FILING DATE: 1998-03-04

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1604

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: CDS

LOCATION: (137)..(1318)

US-09-260-843-1

Query Match 80.0%; Score 16; DB 4; Length 1604;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ggcttgctgggcttg 17

Db 58 GGCTTGCTGGGCTTG 43

RESULT 3

US-08-038-682-6/c

Sequence 6, Application US/08038682

Patent No. 5549897

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: ST. GEME III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/038,682

FILING DATE: 16-MAR-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERKSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-293

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-038-682-6

Query Match

Best Local Similarity 79.0%; Score 15.8; DB 1; Length 9323;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 ggcttgctgggcttgct 20

Db 7071 GGCTTGCTGGGCTTGCTT 7053

RESULT 4

US-08-302-832-6/c

Sequence 6, Application US/08302832

Patent No. 5603938

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,832

FILING DATE: 16-SEP-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US pct/us93/02166

FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-404

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 9323 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-302-832-6

Query Match

Best Local Similarity 79.0%; Score 15.8; DB 1; Length 9323;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgcttgagctgctt 20
|||||
DB 7071 GGCTTGCTTGACTTGCTT 7053

RESULT 5

US-08-530-198-6/c
; Sequence 6, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME ILL, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530.198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-6

Query Match 79.0%; Score 15.8; DB 2; Length 9323;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgcttgagctgctt 20
|||||
DB 7071 GGCTTGCTTGACTTGCTT 7053

RESULT 6
US-08-469-880-6/c
; Sequence 6, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-6

Query Match 79.0%; Score 15.8; DB 2; Length 9323;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgcttgagctgctt 20
|||||
DB 7071 GGCTTGCTTGACTTGCTT 7053

RESULT 7
US-08-728-470-6/c
; Sequence 6, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-728-470-6

Query Match 79.0%; Score 15.8; DB 2; Length 9323;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgctggcctgctt 20
|||||
Db 7071 ggcctgctggcctgctt 7053

RESULT 8
US-08-617-697-6/c
Sequence 6, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-6

Query Match 79.0%; Score 15.8; DB 2; Length 9323;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcctgctggcctgctt 20
|||||
Db 7071 ggcctgctggcctgctt 7053

RESULT 9
US-08-719-641-6/c
Sequence 6, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 9323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-719-641-6

Query Match 79.0%; Score 15.8; DB 4; Length 9323;
Best Local Similarity 89.5%; Pred. No. 69;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ggcttgctggcttgctt 20

Db 7071 GGCTTGCTTGAGTTGCTT 7053

RESULT 10

US-08-390-162-3
; Sequence 3, Application US/08390162
; Patent No. 5576192
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostoglandin E Receptors, Their DNA and
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,162
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024179
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: JP 036580-1992
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 064889-1992
; FILING DATE: 23-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 04221-0020-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1405 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-162-3

Query Match 77.0%; Score 15.4; DB 1; Length 1405;
Best Local Similarity 94.1%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcttgctggcttgctt 19

Db 736 GCCTGCTTGCTTGCTT 752

RESULT 11

US-08-685-945B-3
; Sequence 3, Application US/08685945B
; Patent No. 5804415
; GENERAL INFORMATION:

APPLICANT: Ichikawa, Atsushi

APPLICANT: Narumiya, Shuh

TITLE OF INVENTION: Prostoglandin E Receptors, Their DNA and

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finegan, Henderson, Farabow, Garrett &

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,945B

FILING DATE: 22-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/024179

FILING DATE: 23-FEB-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 036580-1992

FILING DATE: 24-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 064889-1992

FILING DATE: 23-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fordis, Jean B.

REGISTRATION NUMBER: 32,984

REFERENCE/DOCKET NUMBER: 04221-0020-02000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000

TELEFAX: 202-408-4000

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1405 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-685-945B-3

Query Match 77.0%; Score 15.4; DB 1; Length 1405;
Best Local Similarity 94.1%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcttgctggcttgctt 19

Db 736 GCCTGCTTGCTTGCTT 752

RESULT 12

US-08-390-162-1
; Sequence 1, Application US/08390162
; Patent No. 5576192
; GENERAL INFORMATION:
; APPLICANT: Ichikawa, Atsushi
; APPLICANT: Narumiya, Shuh
; TITLE OF INVENTION: Prostoglandin E Receptors, Their DNA and
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390.162
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2107 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-390-162-1

Query Match 77.0%; Score 15.4; DB 1; Length 2107;
Best Local Similarity 94.1%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gcttgccttgagcttgct 19
|||
Db 749 gccttccttgagcttgct 765

RESULT 13
US-08-685-945B-1
Sequence 1, Application US/08685945B
Patent No. 5804415
GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Natumiya, Shuh
TITLE OF INVENTION: prostoglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685.945B
FILING DATE: 22-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2107 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-685-945B-1

Query Match 77.0%; Score 15.4; DB 1; Length 2107;
Best Local Similarity 94.1%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gcttgccttgagcttgct 19
|||
Db 749 gccttccttgagcttgct 765

RESULT 14
US-08-306-691B-24/C
Sequence 24, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
STREET: Two Penn Center, Suite 1800
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306.691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 271 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-306-691B-24

Query Match 74.0%; Score 14.8; DB 1; Length 271;
 Best Local Similarity 88.9%; Pred. No. 1.2e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ggcttgcttgggcttgc 19
 ||| | ||||| |||||
 Db 44 GGCATCCTTGGGCTTGCT 27

RESULT 15

PCT-US93-06251-66/C
 ; Sequence 66, Application PC/TUS9306251
 ; GENERAL INFORMATION:
 ; APPLICANT: Wickstrom, Eric and Rife, Jason P.
 ; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
 ; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
 ; NUMBER OF SEQUENCES: 93
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/06251
 ; FILING DATE: 19930630
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 8586
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 516-742-4343
 ; TELEFAX: 516-742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 271 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; PCT-US93-06251-66

Query Match 74.0%; Score 14.8; DB 5; Length 271;
 Best Local Similarity 88.9%; Pred. No. 1.2e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ggcttgcttgggcttgc 19
 ||| | ||||| |||||
 Db 44 GGCATCCTTGGGCTTGCT 27

Search completed: December 8, 2001, 10:53:20
 Job time: 4990 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 12:14:08 ; Search time 288.04 Seconds
(without alignments)
59.528 Million cell updates/sec

Title: US-09-508-147-23

Perfect score: 20
Sequence: 1 ggagcttgcttgagcttgctt 20

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

N.Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAK75936	Artificial p53 DNA
2	17.4	87.0	342	AAH35155	Human colon cancer
3	17.4	87.0	4034	AAA96706	Reporter gene YPK2
4	16	80.0	732	AAQ21313	DNA encoding envl
5	15.8	79.0	1604	AAZ20440	Maize uroporphyrin
6	15.8	79.0	9323	AAQ49509	Gene cluster for h
7	15.8	79.0	9323	AAK90997	Non-tyrosine Haemo
8	15.4	77.0	901	AAA48936	Corn diacylglycerol
9	15.4	77.0	1200	AAQ94313	Tyrosine phosphata
10	15.4	77.0	1405	AAQ6125	PE2 receptor (EP3
11	15.4	77.0	2107	AAQ6124	PE2 receptor (EP3

12	15.4	77.0	3567	20	AAK13164
13	15.4	77.0	3772	21	AAZ64961
14	15.4	77.0	3772	22	AAK21439
15	15.4	77.0	3772	22	AAK44107
16	15.4	77.0	5000	21	AAK96364
17	15.2	76.0	512	22	AAK68106
18	15.2	76.0	546	22	AAH10683
19	15.2	76.0	588	22	AAH18412
20	15.2	76.0	588	22	AAI28280
21	15.2	76.0	588	22	AAI43528
22	15.2	76.0	588	22	AAI57325
23	15.2	76.0	623	22	AAK22934
24	15.2	76.0	832	22	AAH05512
25	15.2	76.0	928	21	AAK47312
26	15.2	76.0	931	21	AAK37071
27	15.2	76.0	982	22	AAI18356
28	15.2	76.0	982	22	AAI18356
29	15.2	76.0	1107	22	AAI43471
30	15.2	76.0	1244	21	AAK50303
31	15.2	76.0	1247	21	AAK32991
32	15.2	76.0	1299	22	AAI58769
33	15.2	76.0	1359	22	AAI60555
34	15.2	76.0	1480	21	AAZ56772
35	15.2	76.0	1896	21	AAK47249
36	15.2	76.0	2311	22	AAH15918
37	15.2	76.0	2549	22	AAH15872
38	15.2	76.0	2765	21	AAK96707
39	15.2	76.0	3150	21	AAK76991
40	15.2	76.0	9558	16	AAQ88228
41	15.2	76.0	50000	21	AAA96365
42	15	75.0	51	22	AAH79410
43	15	75.0	305	20	AAH7700
44	15	75.0	600	19	AAV4571
45	15	75.0	5832	21	AAK74964

ALIGNMENTS

RESULT 1	
AAK75936	AAK75936 standard; DNA; 20 BP.
ID	AAK75936;
AC	AAK75936;
XX	29-JUL-1999 (first entry)
DT	29-JUL-1999 (first entry)
XX	Artificial p53 DNA fragment SEQ ID NO:23.
DE	Artificial p53 DNA fragment SEQ ID NO:23.
KM	Mouse; p53CP; p53; tumour suppressor; cell growth regulation;
KW	DNA binding site; p53 competing protein; carcinogenesis; apoptosis;
KW	tumour cell growth inhibition; genome guardian; differentiation;
KW	senescence; angiogenesis; ss.
XX	
OS	Synthetic.
OS	Mus musculus.
XX	
PN	W09925820-A1.
XX	
PD	27-MAY-1999.
XX	
PF	10-NOV-1998; 98WO-US23992.
XX	
PR	17-NOV-1997; 97US-0065740.
XX	
PA	(WARNER) WARNER LAMBERT CO.
XX	
PI	Bian J, Sun Y;
XX	
DR	WPI; 1999-347468/29.
XX	
PT	New p53CP protein that specifically binds to the p53 consensus
PT	binding sites, useful for treating p53 associated disorders

XX Claim 4; Page 28; 37pp; English.

XX The present invention describes a p53 competing protein designated p53CP
XX (40 kd) that specifically binds to the p53 consensus binding sites.
XX The p53CP protein is useful for p53 inactivation and regulation during
XX human carcinogenesis, in treat growth arrest and apoptosis, tumour cell
XX growth inhibition, genome guardian, differentiation, senescence, and
XX angiogenesis. The present sequence represents an artificial p53 DNA
XX fragment made from 2 mouse sequences, from the present invention.

XX Sequence 20 BP; 0 A; 4 C; 8 G; 8 T; 0 other;

XX

XX Query Match 100.0%; Score 20; DB 20; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 2,1;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggcctgctggtgctgctt 20
XXXXXXXXXXXXXXXXXXXX
Db 1 ggcctgctggtgctgctt 20

RESULT 2
AAH35155
ID AAH35155 standard; cDNA; 342 BP.

XX AAH35155;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:2237.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
XX P-PSDB; AAG75750.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3741; 9803pp; English.

XX AAH32943 to AAH37195 and AAG75514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

[illegible]

XX Sequence 4034 BP; 1279 A; 778 C; 814 G; 1163 T; 0 other;
SQ

Query Match 87.0%; Score 17.4; DB 21; Length 4034;
Best Local Similarity 94.7%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggccttgctggccttgctt 20
|||||
DB 3053 GGCCTGCTTCGCGCTGCTT 3035

RESULT 4

AAQ21313
ID AAQ21313 standard; DNA; 732 BP.

AC AAQ21313;

DT 01-JUN-1992 (first entry)

DE DNA encoding envelope proteins from filamentous viruses.

XX Garlic species; garlic latent virus (GLV); garlic mosaic virus (GMV);

KW resistance; ss.

XX Allium sativum.

OS
FH Key Location/Qualifiers

FT mat_peptide 4..732
/*tag= a
/product= envelope_protein

PN JP04004879-A.

PD 09-JAN-1992.

PF 18-OCT-1990; 90JP-0277686.

PR 05-MAR-1990; 90JP-0051779.

PR 18-OCT-1989; 89JP-0269118.

PR 18-OCT-1990; 90JP-0277686.

XX (WAKU-) WAKUNAGA SEIYAKU KK.

XX WPI: 1992-060496/08.

DR P-PSDB; AAR20929.

XX Filamentous viral gene - which is used for culture of garlic

PT species resistant to mosaic virus and latent virus

PS Claim 2; Page 2; 22pp; Japanese.

XX The DNA sequence was obtd. by cloning the genes encoding envelope

CC proteins from filamentous viruses. The virus was purified from

CC garlic mosaic virus or garlic latent virus infected garlic leaves.

CC Viral cDNA was prepd. by fractionation of RNA, followed by ligation

CC of polyA and introduction into vectors by transformation. The

CC process can also be performed chemically. Knowledge of the DNA

CC sequence of the filamentous virus envelope proteins allows culturing

CC of new garlic species resistant to GMV and GLV, for the study of

CC viral diseases.

CC See also AAQ21312-9.

XX Sequence 732 BP; 221 A; 200 C; 153 G; 158 T; 0 other;

SQ

DB 286 cttgcttgaggcttgct 301

RESULT 5
AAZ20440/C
ID AAZ20440 standard; DNA; 1604 BP.

XX AAZ20440;

DT 19-NOV-1999 (first entry)

DE Maize uroporphyrinogen decarboxylase coding sequence.

XX Uroporphyrinogen decarboxylase; UROD; maize; uroporphyrinogen III;

KW coproporphyrinogen III; C-5 porphyrin pathway enzyme; sunscreen; therapy;

KW pathogen resistance; disease resistance enhancer; cell death regulator;

KW porphyrin metabolism; photosensitive porphyrin; herbicide resistance;

KW malignant cell growth prevention; ss.

XX Zea mays.

OS
FH Key Location/Qualifiers

FT CDS 137..1318
/*tag= a
/product= UROD

PN W09945125-A2.

PD 10-SEP-1999.

PF 03-MAR-1999; 99WO-US04702.

PR 04-MAR-1998; 98US-0076754.

XX (PION-) PIONEER HI-BRED INT INC.

XX (UMOR) UNIV MISSOURI.

XX Johnal GS, Briggs SP, Gray J, Hu G;

XX WPI: 1999-551045/46.

DR P-PSDB; AAY39471.

XX use of nucleotide sequences encoding enzymes in the C-5 porphyrin

PT pathway for enhancing disease resistance in plants and regulting cell

PT death

PS Claim 3; Page 55-58; 60pp; English.

XX This sequence encodes the maize uroporphyrinogen decarboxylase (UROD) of

CC the invention. The protein is a monooct protein that catalyses the

CC sequential decarboxylation of uroporphyrinogen III to

CC coproporphyrinogen III. The DNA sequence, and other nucleotide sequences

CC encoding enzymes in the C-5 porphyrin pathway can be used for enhancing

CC disease resistance in plants and for regulating cell death. The products

CC can be used for creating or enhancing disease resistance to a pathogen in

CC plants. Expression of the nucleotide sequence in response to pathogenic

CC invasion of a cell effectively disrupts porphyrin metabolism of the

CC transformed plant cell. As a result, photosensitive porphyrins

CC accumulate, leading to a hypersensitive-like response within the invaded

CC cell and development of a localised lesion whereby the spread of the

CC pathogen is contained. The sequences can be used for targeting specific

CC tissues for cell death, e.g. for producing male sterile plants. They can

CC also be used for overcoming herbicide resistance during crop rotation.

CC The antisense nucleotide sequences can also be used in therapies for

CC mammals, particularly humans, for preventing growth of malignant cells.

CC The sequences can also be used in methods for testing the effectiveness

CC of sunscreens.

XX Sequence 1604 BP; 392 A; 365 C; 439 G; 408 T; 0 other;

SQ

Query Match 80.0%; Score 16; DB 20; Length 1604;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cttgcttgaggcttgctt 19

|||||

DB 3053 GGCCTGCTTCGCGCTGCTT 3035

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 ggcttgcttgagcttg 17
|||||
DB 58 GGCTTGCTTGAGCTTG 43

RESULT 6
AA049509/c
ID AA049509 standard; DNA: 9323 BP.
XX
AC AA049509;
XX
DT 26-APR-1994 (first entry)
XX
OS Gene cluster for high molecular weight protein 2 (HmW2).
DE
XX
KW HmW, high molecular weight protein; virus; vaccine; influenza;
KM epitope; immunity; haemophilus influenzae; gene cluster; ss.
XX
OS Haemophilus influenzae.
XX
FH Key Location/Qualifiers
FT CDS 792..5222
FT /tag= a
FT /product= "High molecular weight protein 2"
FT 5375..7009
FT /tag= b
FT /note= "One of a gene cluster for High molecular"
FT weight protein 2"
FT CDS 7149..9098
FT /tag= c
FT /note= "One of a gene cluster for High molecular"
FT weight protein 2"
XX
PN W09319090-A.
XX
PD 30-SEP-1993.
XX
PE 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI
PS Barenkamp SJ;
XX
DR WPI: 1993-320683/40.
DR P-PSDB; AAR41728, AAR41729, AAR41730.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX
PS Claim 10; Figure 7; 100pp; English.
XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
SQ Sequence 9323 BP; 3133 A; 1829 C; 1833 G; 2528 T; 0 other;

Query Match 79.0%; Score 15.8; DB 14; Length 9323;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcttgcttgagcttgctt 20
|||||
DB 7071 GGCTTGCTTGAGTTTGCTT 7053

RESULT 7
AAT90997/c
ID AAT90997 standard; DNA: 9323 BP.
XX
AC AAT90997;
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus influenzae hmw2 gene cluster.
XX
KW Non-typeable Haemophilus; high molecular weight surface protein;
KM HmW2; hmw2 gene; immunogen; vaccine; otitis media; ss.
XX
OS Haemophilus influenzae strain 12.

FH Key Location/Qualifiers
FT CDS 792..5222
FT /tag= a
FT /note= "hmw2 gene"
FT 5375..7009
FT /tag= b
FT /note= "ORF-b"
FT 7249..9198
FT /tag= c
FT /note= "ORF-c"

PD W09736914-A1.
PD 09-OCT-1997.
XX
PF 01-APR-1997; 97WO-US04707.
XX
PR 01-APR-1996; 96US-0617697.
XX
PA (BARE/) BARENKAMP S J.
PI
PS Barenkamp SJ;
XX
DR WPI: 1997-503038/46.
DR P-PSDB; AAM30294.
XX
PT High molecular weight proteins of non-typeable Haemophilus
PT influenzae - useful for vaccine production
XX
PS Disclosure; Page 82-87; 183pp; English.

CC This nucleic acid comprises a gene cluster for the hmw2 gene (see
CC also AAT90997) that encodes high molecular weight surface protein
CC HmW2 (see AAM30294) of non-typeable Haemophilus influenzae strain 12.
CC In addition to the hmw2 gene there are 2 additional downstream
CC open reading frames that are required for the correct processing
CC and secretion of the hmw1 gene product. The ORF-b derived amino
CC acid sequence demonstrates similarity with the derived amino acid
CC sequences of 2 genes which encode proteins required for secretion
CC and activation of haemolysins of P. mirabilis and S. marcescens.
CC HmW proteins (see AAM30291-94) can be used in vaccines, as immunogens
CC for preparation of antibodies and as antigens for detection of
CC these antibodies.
XX
SQ Sequence 9323 BP; 3135 A; 1829 C; 1831 G; 2528 T; 0 other;

Query Match 79.0%; Score 15.8; DB 18; Length 9323;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggcttgcttgagcttgctt 20
|||||
DB 7071 GGCTTGCTTGAGTTTGCTT 7053

RESULT 8

```

AAAA8936/c
ID   AAAA8936 standard; cDNA; 901 BP.
XX
AC   AAAA8936;
XX
DT   06-DEC-2000 (first entry)
XX
DE   Corn diacylglycerol acyltransferase cDNA #4.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KM   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Zea mays.
XX
FH   Key
FT   CDS
FT   Location/Qualifiers
FT       1..858
FT       /tag=a
FT       /partial
FT       /transl_except=(pos:442..444,aa:Xaa)
FT       /transl_except=(pos:490..492,aa:Xaa)
FT       /transl_except=(pos:587..589,aa:Xaa)
FT       /transl_except=(pos:805..807,aa:Xaa)
FT       /transl_except=(pos:820..822,aa:Xaa)
FT       /note="Xaa= unknown"
FT       /product= Partial_diacylglycerol_acyltransferase
FT
XX
PN   MO200032756-A2.
XX
PD   08-JUN-2000.
XX
PE   01-DEC-1999; 99WO-US28354.
XX
PR   02-DEC-1998; 98US-0110602.
PR   31-MAR-1999; 99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR   WPI: 2000-412308/35.
DR   P-PSDB; AAY94516.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 18; Page 46; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is corn diacylglycerol acyltransferase
CC   cDNA from a contig of clones p0042.cspaf49r, p0122.cxmb57r and
CC   p0125.czaa61urb. Diacylglycerol acyltransferases are involved in the
CC   synthesis of triacylglycerols. Alteration of the expression of the
CC   diacylglycerol acyltransferase DNA can be useful for increasing the
CC   level of oils in plant seeds. Inhibitors of diacylglycerol
CC   acyltransferase may be useful as herbicides.
XX
SQ   Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;

```

```

RESULT 9
AAQ94313/c
ID   AAQ94313 standard; cDNA to mRNA; 1200 BP.
XX
AC   AAQ94313;
XX
DT   10-MAY-1996 (first entry)
XX
DE   Tyrosine phosphatase MPP-delta insertion sequence, INS-2.
XX
KW   Tyrosine phosphatase MPP-delta; murine; brain tissue;
KM   glutathione-S-transferase; fusion protein; E. coli; differentiation;
KM   activation; information transmission; nervous system; immune system;
KM   carcinogenesis; insertion; INS-1; INS-2; ds.
XX
OS   Mus musculus domesticus.
XX
PN   JP07236487-A.
XX
PD   12-SEP-1995.
XX
PE   28-FEB-1994; 94JP-0054726.
XX
PR   28-FEB-1994; 94JP-0054726.
XX
PA   (TOKS-) TOKYOTO SHINKAI KAGAKU SOCO KENKYUSHO ZH.
XX
DR   WPI: 1995-347455/45.
DR   P-PSDB; AAR75203.
XX
PT   DNA encoding tyrosine phosphatase MPP-delta - useful for
PT   elucidation of signal transmission mechanisms.
XX
PS   Claim 2; Page 12-13; 14pp; Japanese.
XX
CC   The sequences given in AAQ94312-13 represents insertion sequences which
CC   were included in the tyrosine phosphatase MPP-delta coding sequence.
CC   INS-1 was inserted between the 66th A and the 67th G of the MPP-delta
CC   coding sequence, and INS-2 was inserted between the 1194th A and the
CC   1195th T of the sequence. The MPP-delta sequence was isolated from
CC   murine brain tissue and was cloned, for expression, into the downstream
CC   region of a glutathione-S-transferase sequence and expressed as a fusion
CC   protein in E. coli. MPP-delta proteins regulate differentiation and
CC   activation of cells. This sequence can be used in the elucidation of
CC   the molecular mechanism for information transmission in cells,
CC   the regulation mechanisms in the nervous system or immune system, or in
CC   the mechanism of carcinogenesis.
XX
SQ   Sequence 1200 BP; 345 A; 306 C; 272 G; 277 T; 0 other;

```

Query Match 77.0%; Score 15.4; DB 16; Length 1200;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY   4 ctgtgttggtgctgtt 20
      ||| ||| ||| ||| |||
DB   598 CTAGCTTGGGCTTGCT 582

```

RESULT 10
AAQ46125
ID AAQ46125 standard; cDNA; 1405 BP.
XX
AC AAQ46125;
XX
DT 09-FEB-1994 (first entry)
XX
DE PGE2 receptor (EP3 beta) clone Mp653.
XX
KW prostaglandin E receptor; PGE; agonist; antagonist;
KW GTP-binding protein; digestive tract; constriction; relaxation;

KW	gastricacid; intestinal juice; neurotransmitter; ss.
XX	Mus musculus.
XX	
XX	Key Location/Qualifiers
FH	106..1191
FT	/tag= a
CDS	/product= PGE2_receptor
FT	/note= "PGE subtype EP3 beta"
FT	
XX	
PN	EP557966-A.
PD	01-SEP-1993.
XX	
PE	24-FEB-1993; 93EP-0102873.
XX	
PR	24-FEB-1992; 92JP-0036580.
PR	23-MAR-1992; 92JP-0064889.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
PI	Ichikawa A, Narumiya S;
XX	
DR	WPI: 1993-274435/35.
DR	P-PsDB; AAK42281.
XX	
PT	New prostaglandin E receptor protein and DNA encoding it - used
PT	to study prostaglandin binding, agonists and antagonists etc.
PS	Claim 15; Fig 4; 22pp; English.
XX	
CC	PGE2 is known to be involved in digestive tract constriction/ relaxation, gastric juice secretion and neurotransmitter release. The protein encoded by the gene is capable of receiving PGE and of activating GTP-binding protein. MP653 (AA046125) is identical to MP660 (AA046124) with the exception of a 89-bp deletion in the coding region of the C-terminal tail of the receptor. This deletion creates a new reading frame downstream from this junction. This results in a 30 amino acid C-terminal fragment of the receptor being replaced with a new 26-amino acid fragment in the C-terminal end of the MP653 (AA046125) receptor.
XX	
SO	Sequence 1405 BP; 251 A; 415 C; 418 G; 321 T; 0 other:
Query Match	77.0%; Score 15.4; DB 14; Length 1405;
Best Local Similarity	94.1%; Pred. NO. 3.4e+02;
Matches 16; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	3 gcttgctgggctgtgct 19
Db	736 gccctgctgggctgtgct 752
RESULT 11	
AA046124	
ID	AA046124 standard; cDNA: 2107 BP.
XX	
AC	AA046124;
XX	
DT	09-FEB-1994 (first entry)
XX	
DE	PGE2 receptor (EP3 alpha) clone MP660.
XX	
KW	prostaglandin E receptor; PGE; agonist; antagonist;
KW	GTP-binding protein; digestive tract; constiction; relaxation;
KW	gastric acid; intestinal juice; neurotransmitter; ss.
XX	
OS	Mus musculus.
XX	
FH	Location/Qualifiers
FT	119..1216
CDS	/tag= a
TT	

```

FT      /product= PGE2_receptor
FT      /note= "PGE subtype EP3 alpha"
PN      EP557966-A.
XX
XX      01-SEP-1993.
XX
XX      24-FEB-1993;    93EP-0102873.
XX
XX      24-FEB-1992;    92JP-0036580.
XX      23-MAR-1992;    92JP-0064889.
XX
XX      (TAKE ) TAKEDA CHEM IND LTD.
XX
XX      Ichikawa A, Narumiya S;
XX
XX      WPI; 1993-274435/35.
XX      P-PSDB; AAR42280.
XX
XX      New prostaglandin E receptor protein and DNA encoding it - used
XX      to study prostaglandin binding, agonists and antagonists etc.
XX
XX      Claim 15; Fig 1; 22pp; English.
XX
XX      PGE2 is known to be involved in digestive tract constriction/
XX      relaxation, gastric juice secretion and neurotransmitter release.
XX      The protein encoded by the gene is capable of receiving PGE and of
XX      activating GTP-binding protein.
XX
XX      Sequence 2107 BP; 482 A; 549 C; 563 G; 513 T; 0 other;
SO
OY      3 gctgcttgaggctgct 19
      11111111111111111111
DB      749 gctctgctgggctgct 765

RESULT 12
AAI13164
ID      AAI13164 standard; DNA; 3567 BP.
XX
XX      AAI13164;
AC
XX      19-MAR-1999 (first entry)
XX
XX      Enterococcus faecalis genome contig SEQ ID NO:227.
DE
XX
XX      Enterococcus faecalis; contig; detection: Enterococcal infection;
KW      vaccine; attenuation; computer readable medium; ds.
XX
XX      Enterococcus faecalis.
OS
XX      WO9850555-A2.
XX
XX      12-NOV-1998.
XX
XX      04-MAY-1998;    98WO-US08985.
XX
XX      14-NOV-1997;    97US-0066009.
XX      06-MAY-1997;    97US-0044031.
XX      16-MAY-1997;    97US-0046555.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Barash SC, Dillon PJ, Kunsch CA;
XX
XX      WPI; 1999-045171/04.
XX
XX      New isolated Enterococcus faecalis polynucleotides and polypeptides
XX

```

PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
XX infection.
PS Claim 1, Page 1142-1144; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AXI12938 to AXI13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 3567 BP; 924 A; 617 C; 888 G; 1134 T; 4 other:

Query Match 77.0%; Score 15.4; DB 20; Length 3567;
Best Local Similarity 94.1%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggcgtgtcgtggcgtcgtg 17
|||||
Db 1189 ggcgtgtcgtggcgtcgtg 1205

RESULT 13
AAZ64961
ID AAZ64961 standard; cDNA; 3772 BP.
XX
AC AAZ64961;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO1310 encoding cDNA.
XX
KM Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KM pharmaceutical; receptor immunoadhesin; gene mapping; ss.
OS Homo sapiens.
XX
PN WO963088-A2.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-US12252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.

PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089601.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 23-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 25-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 26-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091526.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.

04-AUG-1998; 980S-0095282.
PR 04-AUG-1998; 980S-0095285.
PR 04-AUG-1998; 980S-0095301.
PR 04-AUG-1998; 980S-0095302.
PR 04-AUG-1998; 980S-0095318.
PR 04-AUG-1998; 980S-0095321.
PR 04-AUG-1998; 980S-0095325.
PR 10-AUG-1998; 980S-0095916.
PR 10-AUG-1998; 980S-0095929.
PR 10-AUG-1998; 980S-0096012.
PR 11-AUG-1998; 980S-0096143.
PR 11-AUG-1998; 980S-0096146.
PR 12-AUG-1998; 980S-0096329.
PR 17-AUG-1998; 980S-0096757.
PR 17-AUG-1998; 980S-0096766.
PR 17-AUG-1998; 980S-0096768.
PR 17-AUG-1998; 980S-0096773.
PR 17-AUG-1998; 980S-0096791.
PR 17-AUG-1998; 980S-0096867.
PR 17-AUG-1998; 980S-0096891.
PR 17-AUG-1998; 980S-0096894.
PR 17-AUG-1998; 980S-0096895.
PR 17-AUG-1998; 980S-0096897.
PR 18-AUG-1998; 980S-0096949.
PR 18-AUG-1998; 980S-0096950.
PR 18-AUG-1998; 980S-0096959.
PR 18-AUG-1998; 980S-0096960.
PR 18-AUG-1998; 980S-0097022.
PR 19-AUG-1998; 980S-0097141.
PR 20-AUG-1998; 980S-0097218.
PR 24-AUG-1998; 980S-0097661.
PR 26-AUG-1998; 980S-0097951.
PR 26-AUG-1998; 980S-0097952.
PR 26-AUG-1998; 980S-0097954.
PR 26-AUG-1998; 980S-0097955.
PR 26-AUG-1998; 980S-0097971.
PR 26-AUG-1998; 980S-0097974.
PR 26-AUG-1998; 980S-0097978.
PR 26-AUG-1998; 980S-0097979.
PR 26-AUG-1998; 980S-0097986.
PR 26-AUG-1998; 980S-0098014.
PR 31-AUG-1998; 980S-0098525.
PR 16-SEP-1998; 980S-0100634.
PR 12-JAN-1999; 990S-0115565.
XX
XX (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR P-PSDB: AAV66645.
XX
XX Membrane-bound proteins and related nucleotide sequences -
XX
XX Claim 2: Fig 35; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
XX polynucleotides encoding them. The PRO sequences of the invention were
XX identified based on extracellular domain homology screening. The PRO
XX sequences have homology with proteins including LDL receptors, TIE
XX ligands and various enzymes. The membrane-bound proteins and receptor
XX molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX immunoadhesins, for instance, can be used as therapeutic agents to block
XX receptor-ligand interactions. The membrane-bound proteins can also be
XX employed for screening of potential peptide or small molecule inhibitors
XX of the relevant receptor/ligand interaction. The PRO encoding sequences
XX are useful as hybridization probes, in chromosome and gene mapping and in
XX the generation of antisense RNA and DNA. PRO nucleic acid sequences
XX will also be useful for the preparation of PRO polypeptides, especially
XX by recombinant techniques.
XX
XX Sequence 3772 BP: 932 A; 1004 C; 1079 G; 757 T; 0 other;

Query Match 77.0%; Score 15.4; DB 21; Length 3772;
Best Local Similarity 94.1%; Pred No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gggcttcttgagcttg 17
||||| |||||||||
Db 3 gggcttcttgagcttg 19
RESULT 14
AAS21439
ID AAS21439 standard; cDNA; 3772 BP.
XX
XX AAS21439;
AC
XX
XX 24-OCT-2001 (first entry)
DT
XX
XX Human cDNA sequence encoding for PRO1310 polypeptide.
DE
XX
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO200140466-A2.
PN
XX
PD 07-JUN-2001.
XX
XX
PF 01-DEC-2000; 2000MO-US32678.
XX
PR 01-DEC-1999; 99MO-US28301.
PR 01-DEC-1999; 99MO-US28634.
PR 02-DEC-1999; 99MO-US28551.
PR 02-DEC-1999; 99MO-US28564.
PR 02-DEC-1999; 99MO-US28565.
PR 09-DEC-1999; 99MO-US170262.
PR 16-DEC-1999; 99MO-US30095.
PR 20-DEC-1999; 99MO-US30911.
PR 20-DEC-1999; 99MO-US30999.
PR 30-DEC-1999; 99MO-US31243.
PR 06-JAN-2000; 2000MO-US00277.
PR 06-JAN-2000; 2000MO-US00376.
PR 11-FEB-2000; 2000MO-US03565.
PR 18-FEB-2000; 2000MO-US04341.
PR 18-FEB-2000; 2000MO-US04342.
PR 22-FEB-2000; 2000MO-US04414.
PR 24-FEB-2000; 2000MO-US04914.
PR 24-FEB-2000; 2000MO-US05004.
PR 01-MAR-2000; 2000MO-US05601.
PR 20-MAR-2000; 2000MO-US07377.
PR 31-MAR-2000; 2000MO-US08439.
PR 17-MAY-2000; 2000MO-US13705.
PR 22-MAY-2000; 2000MO-US14042.
PR 30-MAY-2000; 2000MO-US14941.
PR 02-JUN-2000; 2000MO-US15264.
PR 10-NOV-2000; 2000MO-US30873.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
XX Geritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
XX Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI: 2001-408281/43.
DR P-PSDB: AAU12367.
XX
XX Isolated, secretory and transmembrane PRO polypeptide used to detect
XX other PRO polypeptides, link bioactive molecules to cells expressing

PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical
XX
PS Claim 3; Fig 391; 813pp: English.
CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PMNCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 3772 BP; 932 A; 1004 C; 1079 G; 757 T; 0 other;

Query Match 77.0%; Score 15.4; DB 22; Length 3772;
Best Local Similarity 94.1%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gggtctgttggttggtg 17
||||| |||||||||
Db 3 gggtctcttggttggtg 19

RESULT 15
AAFA4107
ID AAFA4107 standard; cDNA: 3772 BP.
AC AAFA4107;
XX
DT 02-APR-2001 (first entry)
XX
DE Human PRO1310 nucleotide sequence SEQ ID NO:61.
XX
KW Human; secreted and transmembrane protein; PRO; cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay; ss.
XX
OS Homo sapiens.
XX
PM WO200073454-A1.
XX
PD 07-DEC-2000.
XX
PF 30-MAR-2000; 2000WO-US08439.
XX
XX 02-JUN-1999; 99WO-US12252.
XX 23-JUN-1999; 99US-0141037.
XX 07-JUL-1999; 99US-0143048.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 28-JUL-1999; 99US-0146222.
XX 17-AUG-1999; 99US-0149396.
XX 15-SEP-1999; 99WO-US21090.
XX 15-SEP-1999; 99WO-US21547.
XX 08-OCT-1999; 99US-0158663.
XX 30-NOV-1999; 99WO-US28313.
XX 01-DEC-1999; 99WO-US28301.
XX 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
XX
XX (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gebert H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kijavlin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI: 2001-032160/04.
DR P-PSDB; AAB65168.
XX
XX PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
XX
PS Claim 2; Fig 35; 935pp: English.

The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAFA4270 to AAFA4470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAFA4087 to AAFA4269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX

SQ Sequence 3772 BP; 932 A; 1004 C; 1079 G; 757 T; 0 other;

Query Match 77.0%; Score 15.4; DB 22; Length 3772;
Best Local Similarity 94.1%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gggtctgttggttggtg 17
||||| |||||||||
Db 3 gggtctcttggttggtg 19

Search completed: December 8, 2001, 12:14:10
Job time: 9650 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:51:23 ; Search time 1723.37 Seconds
(without alignments)
191.453 Million cell updates/sec

Title: US-09-508-147-23

Perfect score: 20

Sequence: 1 ggagcttgcttgagcttgctt 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.4	92.0	207944	2	AC016999	AC016999 Homo sapi
2	17.4	87.0	2449	8	YSCYKR2A	M24929 Saccharomyc
3	17.4	87.0	35811	8	SC9718	Z49702 S.cerevisia
4	17.4	87.0	50814	2	AC055816	AC055816 Homo sapi
5	17.4	87.0	141036	2	AC023807	AC023807 Mus muscu
6	17.4	87.0	143729	2	AL356353	AL356353 Homo sapi
7	17.4	87.0	154280	2	AC037480	AC037480 Homo sapi
8	17.4	87.0	155808	2	AC018956	AC018956 Homo sapi
9	17.4	87.0	161486	2	AL590008	AL590008 Homo sapi
10	17.4	87.0	166625	2	AC092332	AC092332 Homo sapi
11	17.4	87.0	190460	2	AC024112	AC024112 Mus muscu
12	17.4	87.0	196213	2	AC023948	AC023948 Homo sapi
13	17.4	87.0	201126	2	AL356310	AL356310 Homo sapi
14	17.4	87.0	205932	2	AC073788	AC073788 Mus muscu
15	17.4	87.0	215653	2	AC073742	AC073742 Mus muscu
16	17.4	87.0	217521	2	AC016047	AC016047 Homo sapi
17	17.4	85.0	5741	9	HSIGF2RX3	X83701 Homo sapien
18	17.4	85.0	129968	9	HS249P5	AL035691 Human DNA
19	17.4	85.0	137436	9	AF348209	AF348209 Homo sapi
20	17.4	85.0	159813	2	HS62F24	AL450003 Homo sapi
21	17.4	85.0	167744	2	AL162418	AL162418 Homo sapi
22	17.4	85.0	179793	2	AC069494	AC069494 Homo sapi
23	17.4	85.0	189272	2	AL591905	AL591905 Mus muscu
24	17.4	85.0	238717	5	AB023814	AB023814 Bugeranus
25	16.8	84.0	1143	8	NCF0X2	X80052 N.crassa To
26	16.8	84.0	3401	12	AF206717	AF206717 Shuttle v
27	16.8	84.0	11910	12	HS7511L	AL031675 Human DNA
28	16.8	84.0	89521	2	AL136224	AL136224 Homo sapi
29	16.8	84.0	98663	2	AC012330	AC012330 Homo sapi
30	16.8	84.0	105288	9	AC007708	AC007708 Homo sapi
31	16.8	84.0	106650	9	AF043945	AF043945 Homo sapi
32	16.8	84.0	109865	9	AC090750_1	AL035045 Human DNA
33	16.8	84.0	110000	2	HS905611	U91319 Human chrom
34	16.8	84.0	122469	9	AP000679	AP000679 Homo sapi
35	16.8	84.0	132630	2	AC090427	AC090427 Homo sapi
36	16.8	84.0	135305	2	AC009288	AC009288 Homo sapi
37	16.8	84.0	140876	9	AC008132	AC008132 Homo sapi
38	16.8	84.0	145598	9	AC087106	AC087106 Homo sapi
39	16.8	84.0	148334	2	AP000550	AP000550 Homo sapi
40	16.8	84.0	150724	9	AC023491	AC023491 Homo sapi
41	16.8	84.0	150754	2	AC019241	AC019241 Homo sapi
42	16.8	84.0	150754	2	AP001150	AP001150 Homo sapi
43	16.8	84.0	155804	9	AC013360	AC013360 Homo sapi
44	16.8	84.0	157282	9		
45	16.8	84.0	159550	2		

ALIGNMENTS

RESULT 1
AC016999 207944 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 2 clone RP11-40B20, WORKING DRAFT SEQUENCE,
DEFINITION 19 unordered pieces.
ACCESSION AC016999
VERSION AC016999.5 GI:8954195
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207944)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207944)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington

COMMENT

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8569736.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0040B20
----- Summary Statistics -----
Sequencing vector: M13, 70%
Sequencing vector: plasmid, 30%
Chemistry: Dye-primer ET, 63% of reads
Chemistry: Dye-terminator Big Dye, 37% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 192433 bases at least Q40
Consensus quality: 197366 bases at least Q30
Consensus quality: 200361 bases at least Q20
Insert size: 185000; agarose-fp
Quality coverage: 4.07 in Q20 bases; sum-of-contigs
Quality coverage: 4.11 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1245: contig of 1245 bp in length
* 1246 1345: gap of unknown length
* 1346 2904: contig of 1559 bp in length
* 2905 3004: gap of unknown length
* 3005 5241: contig of 2237 bp in length
* 5242 5341: gap of unknown length
* 5342 6936: contig of 1595 bp in length
* 6937 7036: gap of unknown length
* 7037 9271: contig of 2235 bp in length
* 9272 9371: gap of unknown length
* 9372 11997: contig of 2636 bp in length
* 11998 12097: gap of unknown length
* 12098 16345: contig of 4248 bp in length
* 16346 16445: gap of unknown length
* 16446 20461: contig of 4016 bp in length
* 20462 20561: gap of unknown length
* 20562 25034: contig of 4473 bp in length
* 25035 25134: gap of unknown length
* 25135 31754: contig of 6620 bp in length
* 31755 31854: gap of unknown length
* 31855 38393: contig of 6539 bp in length
* 38394 38493: gap of unknown length
* 38494 45102: contig of 6609 bp in length
* 45103 45202: gap of unknown length
* 45203 52379: contig of 7177 bp in length
* 52380 52479: gap of unknown length
* 52480 60115: contig of 7636 bp in length
* 60116 60215: gap of unknown length
* 60216 72357: contig of 12042 bp in length
* 72358 72357: gap of unknown length
* 72358 92768: contig of 20411 bp in length
* 92769 92868: gap of unknown length
* 92869 119179: contig of 26311 bp in length
* 119180 119279: gap of unknown length
* 119280 156247: contig of 36968 bp in length
* 156248 207944: gap of unknown length
* 156348 207944: contig of 51597 bp in length.

FEATURES

source 1. 207944
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="2"
/clone="RP11-40B20"
1. 1245
/note="assembly_name:Contig9"
1346..2904
/note="assembly_name:Contig10"
3005..5241
/note="assembly_name:Contig11"
5342..6936
/note="assembly_name:Contig12"
7037..9271
/note="assembly_name:Contig13"
9372..11997
/note="assembly_name:Contig14"
12098..16345
/note="assembly_name:Contig15"
16446..20461
/note="assembly_name:Contig16
vector_side:left"
clone_end:r7
20562..25034
/note="assembly_name:Contig17"
25135..31754
/note="assembly_name:Contig18"
31855..38393
/note="assembly_name:Contig19"
38494..45102
/note="assembly_name:Contig20"
45203..52379
/note="assembly_name:Contig21"
52480..60115
/note="assembly_name:Contig22"
60216..72257
/note="assembly_name:Contig23"
72358..92768
/note="assembly_name:Contig24
clone_end:sp6
vector_side:left"
92869..119179
/note="assembly_name:Contig25"
119280..156247
/note="assembly_name:Contig26"
156348..207944
/note="assembly_name:Contig27"
52476 a 49914 c 48857 g 53082 t 1815 others
BASE COUNT
ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 207944;
Best Local Similarity 95.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 99gctgttggtgctgtt 20
DB 11845 GGGCTTGCTGGGCTTCCTT 11864
RESULT 2
YSYCKR2A/C 2449 bp DNA PLN 29-SEP-1995
LOCUS YSYCKR2A Saccharomyces cerevisiae protein kinase (YKR2) gene, complete cds.
DEFINITION M24929.1 GI:295680
ACCESSION M24929.1
VERSION M24929.1
KEYWORDS protein kinase C-related; serine/threonine protein kinase.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 2449)
AUTHORS Kubo, K., Ohno, S., Matsumoto, S., Yahara, I. and Suzuki, K.
TITLE A novel yeast gene coding for a putative protein kinase
JOURNAL Gene 76 (1), 177-180 (1989)
MEDLINE 89306654

FEATURES
source
TATA_signal
gene
CDS

Location/Qualifiers
1..2449
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
78..83
/gene="YKR2"
78..2206
/gene="YKR2"
173..2206
/gene="YKR2"
/codon_start=1
/product="protein kinase"
/protein_id="AA078259.1"
/db_xref="GI:295681"
/translation="MHSWRISKFKLGRKEDGSSDENENKSMGNGLFPHHGEKHND
GSPKRNHEHHIRKINTNETLPSLSLSPKLRNDASFKNPISGINDSKASERKASQ
SSTEQGSSSGLTAVKAVYSGKDTLPFTITSNTIIQKLSSGILTSNDSEVA
AIMROLPRYKRVDDSGEGLIDRAFAKFIIPSSILPSTNSSPLLYFTIEPNSIT
TISPDMGTMEQPVFNKISTPDYTRKRLKIDVAFARIPSLPSPKMMQOEIGDEEVL
KEILKINTNODIHLSDFHLP.LPNLKDISAQOIRLYNHMWSILERGYGLNTIDVKKPS
KNKPLSIDDPFLAVIGKSGFKVQVAKKTKQIKYALKAKRAIYKCEVTHLAE
RTVLRVDCPTIVPLKESFQSEKLYLVLAFTNGELFYHLQHBGRSLARSFYIAE
LLCALDSLAKLIDVYIRDLKPNILLDYOGHIALCDFGLCKLNMKDNKTDITFCGTPY
LAPETLLGQYTKVDMWTIGILLYEMTGLPYDENVPYMKKILQOPLLPDGGFD
PAAKDLIGLSRDSRRLGVNGTDEIRNHFEDISWKKLLKGYLPPYKPIYKSEI
DTANPDEFTKEKPIDSVDEYLSASIOKQGGWTYIGDEQLGDSPSGGRIS"

BASE COUNT 788 a 480 c 497 g 684 t

ORIGIN

Query Match 87.0% Score 17.4; DB 84; Length 2449;
Best Local Similarity 94.7% Pred. No. 2, 1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ggcctgcttgaggctgctt 20
||||||| |||||
Db 2225 GCGTCGCTCGGCTGCTT 2207

RESULT 3
SC9718 35811 bp DNA PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome XIII cosmid 9718.
DEFINITION 249702.271257
ACCESSION 249702.1 GI:817859
VERSION 249702.1
KEYWORDS acetolactate synthase; aldehyde dehydrogenase; beta-transducin;
coiled coil protein; folypolyglutamate synthase; GAL5; glycoseyl
hydroxylase; guanine nucleotide binding protein; ILV2; membrane
protein; myosin homologue; PGM2; phosphoglucomutase; protein
kinase; SMR1; succinate dehydrogenase; transfer RNA-Leu; transfer
RNA-Lys; YKR2; YPK2.
baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 35811)
AUTHORS Hunt,S. and Bowman,S.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 35811)
AUTHORS Barrell,B. and Raftery,M.A.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1995) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1HQ E-mail: barrell@sanger.ac.uk

Notes:
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)

FEATURES
source
CDS
CDS
CDS
CDS

Location/Qualifiers
1..35811
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 9718"
/map="13R"
complement(1..1537)
/note="Ym9718.01c, unknown, len: >512, CAI: 0.14, similar
to SW:YKML_YEAST_P32330 hypothetical Trp-Asp repeats
containing protein in YPK1-SRP21 intergenic region
(YK1121W or YK1525) (49.1% identity in 426 aa overlap)"
/codon_start=1
/product="unknown"
/protein_id="CA089738.1"
/db_xref="GI:817860"
/db_xref="SWISS-PROT:Q03177"
/translation="MTNMAEDKTSISKVHSEFSSNNTTGSEERSHQESFDESSS
GESKSKSLNLEYADIEPLKFMRTNTNMDLKYSNNGNATDSFMRLEHLOGNT
LNSLRYNEFYEPNSIDTEOENLIREPKYIKMLRKKNLKOFRLFLAOLEMVEGE
TVTSTKSSSEPTSKAISTKRSRQKFKMATGSKGKIRIMKVISPVRAFLDSSAES
NKEARAKSMRITKQVSSLSNNPKRFIDSAIEKYEKKLNLVAPVHPPLRIKYE
HVQVDLWMSKNNFLISASMDKTVKLMHPKRNKSLTFIHDFVTCVEHPDTRF
ISGCLDHKRCRLMSILDEVSFEYDCODLITSVTLSPGKATYIIGTFNGVHILMTG
LTPVSSPHVADROTEONAHAVMETEDSKIHGCRVTLQAFRSQDLNDSFRLVYTSND
SRIRFLDEQRKLLLEVLFKGFHSGSSQHKQAQSLIMHGOPIVNSDDHVVGMRLKSSD
REN"

complement(2087..2449)
/note="Ym9718.02c, questionable orf, len: 120, CAI: 0.07"
/codon_start=1
/product="unknown"
/protein_id="CA089739.1"
/db_xref="GI:817861"
/db_xref="SWISS-PROT:Q04436"
/translation="WRRGCALLATVTLRLMLLHFYYSIIRILPSTELRSKTDLLSA
EGRRLIPADVAKKPAKGRKSRLEPFAKRORTLALPERKDSLCLPLPLFASIGRQLIS
SADDPGFCAGSANGSLT"
complement(2605..4638)
/gene="YPK2"
complement(2605..4638)
/gene="YPK2"
/note="Ym9718.03c, YPK2 gene (YKR2), len: 677, CAI: 0.13,
SW:YPK2_YEAST_P18961, serine/threonine-protein kinase
YPK2/YKR2"
/codon_start=1
/product="Ypk2p"
/protein_id="CA089740.1"
/db_xref="GI:817862"
/db_xref="SWISS-PROT:P18961"
/translation="MHSWRISKFKLGRKEDGSSDENENKSMGNGLFPHHGEKHND
GSPKRNHEHHIRKINTNETLPSLSLSPKLRNDASFKNPISGINDSKASERKASQ
SSTEQGSSSGLTAVKAVYSGKDTLPFTITSNTIIQKLSSGILTSNDSEVA
AIMROLPRYKRVDDSGEGLIDRAFAKFIIPSSILPSTNSSPLLYFTIEPNSIT
TISPDMGTMEQPVFNKISTPDYTRKRLKIDVAFARIPSLPSPKMMQOEIGDEEVL
KEILKINTNODIHLSDFHLP.LPNLKDISAQOIRLYNHMWSILERGYGLNTIDVKKPS
KNKPLSIDDPFLAVIGKSGFKVQVAKKTKQIKYALKAKRAIYKCEVTHLAE
RTVLRVDCPTIVPLKESFQSEKLYLVLAFTNGELFYHLQHBGRSLARSFYIAE
LLCALDSLAKLIDVYIRDLKPNILLDYOGHIALCDFGLCKLNMKDNKTDITFCGTPY
LAPETLLGQYTKVDMWTIGILLYEMTGLPYDENVPYMKKILQOPLLPDGGFD
PAAKDLIGLSRDSRRLGVNGTDEIRNHFEDISWKKLLKGYLPPYKPIYKSEI
DTANPDEFTKEKPIDSVDEYLSASIOKQGGWTYIGDEQLGDSPSGGRIS"
complement(5082..6791)
/gene="PGM2"
complement(5082..6791)
/gene="PGM2"
/note="Ym9718.04c, PGM2 gene, (GAL5) SW:PGM2_YEAST_P37012
phosphoglucomutase 2, Ps00710 Phosphoglucomutase and
phosphomannomutase phosphoserine signature"

```
/codon_start=1
/product="Pgm2p"
/protein_id="CA89741.1"
/db_xref="GI:817863"
/db_xref="SWISS-PROT:P37012"
/translation="MSFOETVPTKRPYPEDOKPRTSGLKKTKTKVREDEPNTENTPQSI
MEALDEGSKATLVVGGGGRYNDVILKIAIGANCIKLVIGQHLSTPASHI
MRYEEKCTGIIILASHNPGSPENDMGIKNLGSGPAPSPVNAWEISKLTSTK
IKIDPELDLTIGKNKKYGPLVDIIDITADYVNFLEIFDILKIFKFINOSRSTK
WKLFLDSMNGVTPGYKAIPEVDEGLPADVLOMMHPSPDFGMMHPDENLKYASSLVK
RVREKIEFGAASDGDGRNMITYGSPFSVPSGSVAIAEYAAEIPKARQYIGLA
RSPFSGAIDVAAHGLNCEYVPGWKFPCALDAAKLSCESESESGNNHREKD
GVAMAMANTLAIYNKHHPENESIKITONEFMAKTORTFTRIDEKVELEKANKI
VDOLKATYKSGVNSAFPADESLKVTDCGSDTYTDLGVSVDHGLVTKLSGARFY
LRISGTSSGATIRLYIEKYCDKSOYOKTAEVLEKPLINSIVIKFLEKQVLTGEPT
VR"
```

misc_feature

```
complement(6411..6455)
/gene="PGM2"
/note="P500710 Phosphoglucumutase and phosphomannomutase
phosphoserine signature"
complement(7486..9375)
/note="YM9718.05c, unknown, len: 629, CAI: 0.13"
/codon_start=1
/product="unknown"
/protein_id="CA89742.1"
/db_xref="GI:817864"
/db_xref="SWISS-PROT:Q04437"
/translation="MSESETFIVDVPSPSMKNNNNVSKSMAYLEYTLTKSKSRKTD
MISCYLANCPVSENOEIPNVEFOISLPATVTTATGIFIRLKOYCOOHSDNNG
LOSMIOCLLYSLDICOOFARKILKOIVATVTDLDITDEIDLTETLSRITIL
IDCKRDQDEKKSNMVLVAIPNRIRYNNELLVETSPATSVKRVKRVSELR
GADILSTOTSNPSSGSMODENCLIKVEAFPRKAVSGLNRTAAVEDSOQKERYVG
KSIEYEIHNENGNKNVSEDDSGSYIPVLSDSYTKAVRGADVAVLPVLVDQF
VYESFPGDLDFLNLREALPRYFLTSESSFTADTRLCQDLAASLVDMLENK
IAVARYSKDSEVMCALCPVLIEHNSINEKEFVSLTCLPPEDEGTEVTPKL
LDRTTSGVPLKEEDHIDELMOFVDSMDTDEPIGNYOPDEGVTPTTLP
LPSLNKQDEKKKDLRPTPYVYRQOVLLEHMHQIMINDSREFEIPELPDSLNKI
SPYTHKFTDKLVEVLGIRKVKDLIDSELTLEKRLPDLFLTKRGSHRSGSP
NNSNN"
```

CDS

```
complement(9806..9878)
/note="tRNA-Lys gene, anticodon, ctt, len: 73"
complement(10399..10730)
/note="solo delta element"
10756..11092
/note="solo delta element"
12199..12546
/note="YM9718.06, unknown, len: 115, CAI: 0.12"
/codon_start=1
/product="unknown"
/protein_id="CA89743.1"
/db_xref="GI:817865"
/db_xref="SWISS-PROT:Q04438"
/translation="MGSFMDAFAYDKKHAHDPVYGVGNHNNHNGDSTQVWFSKEYHQ
PRTHOENTLOSMRSSISGSDSDVEDYKBEKRLPAEVEIPRNVDISNMSQCEFLRLYE
SLRGEPDNKKYNR"
13269..13532
/gene="ILV2"
13269..13532
/gene="ILV2"
/note="YM9718.07, ILV2 gene (SMR1) SW:ILVB_YEAST_P07342
acetylactate synthase; P500187 Thiamine pyrophosphate
enzymes signature"
/codon_start=1
/product="Ilv2p"
/protein_id="CA89744.1"
/db_xref="GI:817866"
/db_xref="SWISS-PROT:P07342"
/translation="MIRQSTLKNFAIRCFQHLAARYNTPAMRSVALAQRYSRRY
SASPLPASKRPEPASFNVDPLEQAPASKLAKRAEPMDTSFVGLTGQIFENEMK
SRQNVDFEGYPGGAILPYVDALHNSDKFNFVLPKHEOGAGHAEVARGASGPGVVL
VTSQPGATVYVTPKADAFADGIPMVYFTGVGPSAIGTDAFOADAVYGISRSTCKMY
MKSVEELPLRINAFPIATISGRGPGVLPVDPKDVYRAILRNPLPTKTTIPSNALNO
TISRQDEFVQOSINKADLLINLAKKPVLYGAGILNHADEGRLLKELTDAIIPVTTT
```

gene

CDS

```
LOGLSFQDEDPKSLDMLGMHCATLANLAVQADLLIAGAREDDRVATNISKFPAPA
RRAAEGRGIIIEVSPSKINKRVQOIAVEGATNTLKMMSKIFPVERSEMFQO
INKMKREYPIAMEETPGSKIKPQTVAIKKSVANDGRHVIYTVGQHQMAAOHM
TWRNPHEITISGSGTGWYGLPAHGAOVAKPESLVIDIDGDSFNNLTRELSAVQ
GTPVKIILNNEBGWYOMOSLEHRSHTHOLNDFIKIAEMAKLKGIRYKQDE
LDAKLKEFVSFKGVLLLEVEDKKVPLPMAVAGSGGLDEFINDFEVRQOTELRHR
TGRKH"
14865..14924
/gene="ILV2"
/note="P500187 Thiamine pyrophosphate enzymes signature"
15772..19431
/note="YM9718.08, similar to myosin, len: 1219, CAI: 0.20,
similar to SW:YKM9_YEAST_P36006 myosin heavy chain
homologue YKL129C (75.9% identity in 1090 aa overlap);
P500017 ATP/GTP-binding site motif A; P500176 Eukaryotic
DNA topoisomerase I active site"
/codon_start=1
/product="unknown"
/protein_id="CA89745.1"
/db_xref="GI:817867"
/db_xref="SWISS-PROT:Q04439"
/translation="MALTKRGARKKHVQEPAKRSANIKKATPDSKKKEVGSDDLTL
```

misc_feature

CDS

Query Match 87.0% Score 17.4; DB 8; Length 35011;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gacttgcttgagcttgctt 20
||||| ||||| ||||| |||||

Db 2586 GCCTTGCTTCGGCTTGCTT 2604

RESULT 4
AC055816/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-275H19 map 1, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC055816
VERSION AC055816.1 GI:7582649
KEYWORDS HTG; HTGS_PHRASEO.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 50814)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-275H19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 50814)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bouslavsky,L., Bouhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deareliano,K., Dewar,K., Dietz,J.S.,
Dodg,S., Domilo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McErdan,R., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teague,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (18-Apr-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8009

Center clone name: 275_H_19

* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 647: contig of 647 bp in length
* 648 747: gap of 100 bp
* 748 1335: contig of 588 bp in length
* 1336 1435: gap of 100 bp
* 1436 2130: contig of 695 bp in length
* 2131 2230: gap of 100 bp
* 2231 2813: contig of 583 bp in length
* 2814 2913: gap of 100 bp
* 2914 3582: contig of 669 bp in length
* 3583 3682: gap of 100 bp
* 3683 4338: contig of 656 bp in length
* 4339 4438: gap of 100 bp
* 4439 5091: contig of 651 bp in length
* 5092 5191: gap of 100 bp
* 5192 5794: contig of 603 bp in length
* 5795 5894: gap of 100 bp
* 5895 6508: contig of 614 bp in length
* 6509 6608: gap of 100 bp
* 6609 7262: contig of 654 bp in length
* 7263 7362: gap of 100 bp
* 7363 8039: contig of 677 bp in length
* 8040 8139: gap of 100 bp
* 8140 8792: contig of 653 bp in length
* 8793 8892: gap of 100 bp
* 8893 9595: contig of 703 bp in length
* 9596 9695: gap of 100 bp
* 9696 10337: contig of 642 bp in length
* 10338 10437: gap of 100 bp
* 10438 11089: contig of 652 bp in length
* 11090 11189: gap of 100 bp
* 11190 11817: contig of 628 bp in length
* 11818 11917: gap of 100 bp
* 11918 12528: contig of 611 bp in length
* 12529 12628: gap of 100 bp
* 12629 13286: contig of 656 bp in length
* 13287 13386: gap of 100 bp
* 13387 14049: contig of 663 bp in length
* 14050 14149: gap of 100 bp
* 14150 14784: contig of 635 bp in length
* 14785 14884: gap of 100 bp
* 14885 15548: contig of 664 bp in length
* 15549 15648: gap of 100 bp
* 15649 16248: contig of 600 bp in length
* 16249 16348: gap of 100 bp
* 16349 16904: contig of 556 bp in length
* 16905 17004: gap of 100 bp
* 17005 17585: contig of 581 bp in length
* 17586 17685: gap of 100 bp
* 17686 18312: contig of 627 bp in length
* 18313 18412: gap of 100 bp

18413 19002: contig of 590 bp in length
* 19003 19102: gap of 100 bp
* 19103 19754: contig of 653 bp in length
* 19755 19854: gap of 100 bp
* 19855 20419: contig of 565 bp in length
* 20420 20519: gap of 100 bp
* 20520 21082: contig of 563 bp in length
* 21083 21182: gap of 100 bp
* 21183 21857: contig of 675 bp in length
* 21858 21957: gap of 100 bp
* 21958 22608: contig of 651 bp in length
* 22609 22708: gap of 100 bp
* 22709 23449: contig of 741 bp in length
* 23450 23549: gap of 100 bp
* 23550 24243: contig of 694 bp in length
* 24244 24343: gap of 100 bp
* 24344 25018: contig of 675 bp in length
* 25019 25118: gap of 100 bp
* 25119 25799: contig of 681 bp in length
* 25800 25899: gap of 100 bp
* 25900 26521: contig of 622 bp in length
* 26522 26621: gap of 100 bp
* 26622 27270: contig of 649 bp in length
* 27271 27370: gap of 100 bp
* 27371 28038: contig of 668 bp in length
* 28039 28138: gap of 100 bp
* 28139 28646: contig of 508 bp in length
* 28647 28746: gap of 100 bp
* 28747 29359: contig of 613 bp in length
* 29360 29459: gap of 100 bp
* 29460 30117: contig of 658 bp in length
* 30118 30217: gap of 100 bp
* 30218 30869: contig of 652 bp in length
* 30870 30969: gap of 100 bp
* 30970 31560: contig of 591 bp in length
* 31561 31660: gap of 100 bp
* 31661 32284: contig of 624 bp in length
* 32285 32384: gap of 100 bp
* 32385 33031: contig of 647 bp in length
* 33032 33131: gap of 100 bp
* 33132 33761: contig of 630 bp in length
* 33762 33861: gap of 100 bp
* 33862 34484: contig of 622 bp in length
* 34485 34584: gap of 100 bp
* 34585 35183: contig of 599 bp in length
* 35184 35283: gap of 100 bp
* 35284 35779: contig of 496 bp in length
* 35780 35879: gap of 100 bp
* 35880 36568: contig of 689 bp in length
* 36569 36668: gap of 100 bp
* 36669 37322: contig of 654 bp in length
* 37323 37422: gap of 100 bp
* 37423 38076: contig of 654 bp in length
* 38077 38176: gap of 100 bp
* 38177 38823: contig of 647 bp in length
* 38824 38923: gap of 100 bp
* 38924 39590: contig of 667 bp in length
* 39591 39690: gap of 100 bp
* 39691 40288: contig of 596 bp in length
* 40289 40388: gap of 100 bp
* 40389 41008: contig of 620 bp in length
* 41009 41108: gap of 100 bp
* 41109 41777: contig of 669 bp in length
* 41778 41877: gap of 100 bp
* 41878 42573: contig of 696 bp in length
* 42574 42673: gap of 100 bp
* 42674 43369: contig of 596 bp in length
* 43270 43369: gap of 100 bp
* 43370 44030: contig of 661 bp in length
* 44031 44130: gap of 100 bp
* 44131 44721: contig of 591 bp in length
* 44722 44821: gap of 100 bp
* 44822 45569: contig of 748 bp in length

45570 45669: gap of 100 bp
* 45670 46278: contig of 609 bp in length
* 46279 46378: gap of 100 bp
* 46379 47009: contig of 631 bp in length
* 47010 47109: gap of 100 bp
* 47110 47755: contig of 646 bp in length
* 47756 47855: gap of 100 bp
* 47856 48527: contig of 672 bp in length
* 48528 48627: gap of 100 bp
* 48628 49323: contig of 696 bp in length
* 49324 49423: gap of 100 bp
* 49424 50042: contig of 619 bp in length
* 50043 50142: gap of 100 bp
* 50143 50814: contig of 672 bp in length.
Location/Qualifiers

Query Match 87.0% Score 17.4; DB 2; Length 50814;
Best Local Similarity 94.7%; Pred No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 99gctgcttgagctgct 19
|||||
Db 20041 GGGCGCTGTGGGCTTGC 20023

RESULT 5
AC023807/C
LOCUS AC023807 141036 bp DNA HTG 04-NOV-2000
DEFINITION Mus musculus clone RP23-276G11. *** SEQUENCING IN PROGRESS ***
AC023807
AC023807.5 GI:11079369
VERSION AC023807.5
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 141036)
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okunou, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J., H., Gunaratne, P., Haller, G., Hernandez, J., Hogue, M.,
Korak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 141036)
Worley, K.C.
Direct Submission
Submitted (18-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced gi:8248600.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: MABN
Center clone name: RP23-276G11

Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-Primer Bodypy; 93% of reads

Chemistry: Dye-terminator Big Dye 7% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 89000 bases at least Q40
Consensus quality: 117703 bases at least Q30
Estimated insert size: 129739 bases at least Q20
Estimated insert size: 114732; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently
consists of 60 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
6743 6742: contig of 6742 bp in length
* 6843 6842: gap of unknown length
* 14985 14985: contig of 8143 bp in length
* 14986 15085: gap of unknown length
* 15086 22169: contig of 7084 bp in length
* 22170 22269: gap of unknown length
* 22270 26605: contig of 4336 bp in length
* 26606 26705: gap of unknown length
* 26706 30459: contig of 3754 bp in length
* 30460 30559: gap of unknown length
* 30560 35347: contig of 4788 bp in length
* 35348 35447: gap of unknown length
* 35448 36609: contig of 4162 bp in length
* 36610 39709: gap of unknown length
* 39710 42947: contig of 3238 bp in length
* 42948 43047: gap of unknown length
* 43048 47283: contig of 4236 bp in length
* 47284 47383: gap of unknown length
* 47384 49531: contig of 2148 bp in length
* 49532 52634: gap of unknown length
* 49632 52634: contig of 3003 bp in length
* 52635 52734: gap of unknown length
* 52735 55748: contig of 3014 bp in length
* 55749 55848: gap of unknown length
* 55849 58508: contig of 2660 bp in length
* 58509 61384: gap of unknown length
* 61385 61484: gap of unknown length
* 61485 63452: contig of 1968 bp in length
* 63453 63552: gap of unknown length
* 63553 65016: contig of 1464 bp in length
* 65017 65116: gap of unknown length
* 65117 67313: contig of 2197 bp in length
* 67314 67413: gap of unknown length
* 67414 68393: contig of 1980 bp in length
* 68394 69493: gap of unknown length
* 69494 70964: contig of 1471 bp in length
* 70965 71064: gap of unknown length
* 71065 73551: contig of 2487 bp in length
* 73552 73651: gap of unknown length
* 73652 76297: contig of 2646 bp in length
* 76298 76397: gap of unknown length
* 76398 78996: contig of 2599 bp in length
* 78997 80661: gap of unknown length
* 80662 80761: contig of 1565 bp in length
* 80762 82062: gap of unknown length
* 82063 82162: gap of unknown length
* 82163 83932: contig of 1770 bp in length
* 83933 84032: gap of unknown length
* 84033 87019: contig of 2987 bp in length
* 87020 87119: gap of unknown length
* 87120 90366: contig of 3247 bp in length
* 90367 90466: gap of unknown length


```

* 90467 92356: contig of 1890 bp in length
* 92357 92456: gap of unknown length
* 92457 94598: contig of 2242 bp in length
* 94699 94798: gap of unknown length
* 94799 96484: contig of 1686 bp in length
* 96485 96584: gap of unknown length
* 96585 96655: contig of 2071 bp in length
* 96656 98755: gap of unknown length
* 98756 100564: contig of 1809 bp in length
* 100565 100664: gap of unknown length
* 100665 102615: contig of 1951 bp in length
* 102616 102715: gap of unknown length
* 102716 103729: contig of 1014 bp in length
* 103730 103829: gap of unknown length
* 103830 105670: contig of 1841 bp in length
* 105671 105770: gap of unknown length
* 105771 106991: contig of 1221 bp in length
* 106992 107091: gap of unknown length
* 107092 109177: contig of 2086 bp in length
* 109178 109277: gap of unknown length
* 109278 110785: contig of 1508 bp in length
* 110786 110885: gap of unknown length
* 110886 112715: contig of 1830 bp in length
* 112716 112815: gap of unknown length
* 112816 113848: contig of 1033 bp in length
* 113849 113948: gap of unknown length
* 113949 114998: contig of 1050 bp in length
* 114999 115098: gap of unknown length
* 115099 116175: contig of 1077 bp in length
* 116176 116275: gap of unknown length
* 116276 118035: contig of 1760 bp in length
* 118036 118135: gap of unknown length
* 118136 119198: contig of 1063 bp in length
* 119199 119298: gap of unknown length
* 119299 120798: contig of 1500 bp in length
* 120799 120898: gap of unknown length
* 120899 122238: contig of 1340 bp in length
* 122239 122339: gap of unknown length
* 122339 123486: contig of 1148 bp in length
* 123487 123586: gap of unknown length
* 123587 124888: contig of 1302 bp in length
* 124889 124988: gap of unknown length
* 124989 126483: contig of 1495 bp in length
* 126484 126583: gap of unknown length
* 126584 127935: contig of 1352 bp in length
* 127936 128035: gap of unknown length
* 128036 129167: contig of 1132 bp in length
* 129168 129267: gap of unknown length
* 129268 130323: contig of 1056 bp in length
* 130324 130423: gap of unknown length
* 130424 131722: contig of 1299 bp in length
* 131723 131822: gap of unknown length
* 131823 133063: contig of 1241 bp in length
* 133064 133163: gap of unknown length
* 133164 134665: contig of 1502 bp in length
* 134666 134765: gap of unknown length
* 134766 135826: contig of 1061 bp in length
* 135827 135926: gap of unknown length
* 135927 136953: contig of 1027 bp in length
* 136954 137053: gap of unknown length
* 137054 138375: contig of 1322 bp in length
* 138376 138475: gap of unknown length
* 138476 139538: contig of 1063 bp in length
* 139539 139638: gap of unknown length
* 139639 141036: contig of 1398 bp in length.

```

FEATURES

source

1. 141036
/organism="Mus musculus"
/db_xref="taxon:10090"

/clone="RP23-276G11"

BASE COUNT 37581 a 29162 c 29145 g 39216 t 5932 others
ORIGIN

```

Query Match      87.0%; Score 17.4; DB 2; Length 141036;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      1 99gcctctggcctctgc 19
          ||| ||| ||| ||| |||
Db 55943 GGGCTTGGCTTGACTTGC 55925

```

RESULT 6

AL356353

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

AL356353 143729 bp DNA HTG 20-JAN-2001
Homo sapiens chromosome 1 clone RP11-359B8, *** SEQUENCING IN
PROGRESS ***; 17 unordered pieces.
AL356353
AL356353.7 GI:9988366
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 143729)
Direct Submission
Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Sep 6, 2000 this sequence version replaced g1:9909036.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba359e8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 136131 bases at least Q40
Consensus quality: 138728 bases at least Q30
Consensus quality: 140345 bases at least Q20
Insert size: 142129; sum-of-contigs
Insert size: 172356; 1.1% error; agarose-fp
Quality coverage: 4.42x in Q20 bases; sum-of-contigs Quality
coverage: 4.06x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4513: contig of 4513 bp in length
* 4514 4613: gap of 100 bp
* 4614 11323: contig of 6710 bp in length
* 11324 11423: gap of 100 bp
* 11424 15508: contig of 4085 bp in length
* 15509 15608: gap of 100 bp
* 15609 18097: contig of 2489 bp in length
* 18098 18197: gap of 100 bp
* 18198 53112: contig of 34915 bp in length
* 53113 53212: gap of 100 bp
* 53213 83298: contig of 30086 bp in length
* 83299 83398: gap of 100 bp
* 83399 86477: contig of 3079 bp in length
* 86478 86577: gap of 100 bp
* 86578 89084: contig of 2507 bp in length
* 89085 89184: gap of 100 bp
* 89185 110408: contig of 21224 bp in length

```

```

* 110409 110508: gap of 100 bp
* 110509 112920: contig of 2412 bp in length
* 112921 113020: gap of 100 bp
* 116331 116331: contig of 3311 bp in length
* 116332 116431: gap of 100 bp
* 116432 122305: contig of 5874 bp in length
* 122306 122405: gap of 100 bp
* 122406 125046: contig of 2641 bp in length
* 125047 125146: gap of 100 bp
* 125147 130304: contig of 5158 bp in length
* 130305 130404: gap of 100 bp
* 130405 136103: contig of 5699 bp in length
* 136104 136203: gap of 100 bp
* 136204 140722: contig of 4519 bp in length
* 140723 140822: gap of 100 bp
* 140823 143729: contig of 2907 bp in length.

```

FEATURES

```

SOURCE
1. 143729
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="1"
   /clone="RP11-359E8"
   /clone.lib="RPC1-11.2"
1. 4513
   /note="assembly_fragment:00117"
   /fragment_chain:1"
4614. 111323
   /note="assembly_fragment:01848"
   /fragment_chain:1"
11424. 15508
   /note="assembly_fragment:00385"
   /fragment_chain:2"
15609. 18097
   /note="assembly_fragment:01396"
   /fragment_chain:2"
18198. 53112
   /note="assembly_fragment:00860"
   /fragment_chain:3"
53213. 83298
   /note="assembly_fragment:01688"
   /fragment_chain:3"
83399. 86477
   /note="assembly_fragment:00739"
   /note="assembly_fragment:00809"
86578. 89084
   /note="assembly_fragment:00809"
89185. 110408
   /note="assembly_fragment:00876"
110509. 112920
   /note="assembly_fragment:01075"
113021. 116331
   /note="assembly_fragment:01159"
116432. 122305
   /note="assembly_fragment:01399"
122406. 125046
   /note="assembly_fragment:01505"
125147. 130304
   /note="assembly_fragment:01563"
130405. 136103
   /note="assembly_fragment:01729"
136204. 140722
   /note="assembly_fragment:01876"
140823. 143729
   /note="assembly_fragment:01450"
   /note="assembly_fragment:01450"
   /vector_side:right"
   /vector_end:SP6

```

```

BASE COUNT 37978 a 33378 c 32618 g 38143 t 1612 others
ORIGIN

```

```

Query Match 87.0%: Score 17.4; DB 2; Length 143729;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 1ggcgtcttgaggcttgc 19
Db 40608 GGGCCTGCTGGGCTTGCT 40626

```

```

RESULT 7
AC037480/c
LOCUS
DEFINITION
AC037480 Homo sapiens chromosome 18 clone RP11-63508 map 18, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC037480
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 154280)
Homo sapiens chromosome 18, clone RP11-63508
2 (bases 1 to 154280)
Unpublished
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgatter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menes, L., Mihova, T., Miranda, C., Mlangi, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Titrilli, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zannoun, J., Zimmer, A. and Zody, M.

```

```

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L9289
Center clone name: 635_O-8

```

```

----- Summary Statistics
Sequencing vector: M13: M77815; 3% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150824 bases at least Q40
Consensus quality: 152392 bases at least Q30
Consensus quality: 152924 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 153380; sum-of-contrigs
Quality coverage: 8.6 in Q20 bases; agarose-fp
Quality coverage: 8.2 in Q20 bas.

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```
1 19698: contig of 19698 bp in length
* 19699 19798: gap of 100 bp
* 19799 21284: contig of 1486 bp in length
* 21285 21384: gap of 100 bp
* 21385 25070: contig of 3686 bp in length
* 25071 25170: gap of 100 bp
* 25171 32681: contig of 7511 bp in length
* 32682 32781: gap of 100 bp
* 32782 41653: contig of 8872 bp in length
* 41654 41753: gap of 100 bp
* 41754 45817: contig of 4064 bp in length
* 45818 45917: gap of 100 bp
* 45918 52511: contig of 6594 bp in length
* 52512 52611: gap of 100 bp
* 52612 64034: contig of 11423 bp in length
* 64035 64134: gap of 100 bp
* 64135 135344: contig of 71210 bp in length
* 135345 135445: gap of 100 bp
* 135445 154280: contig of 18836 bp in length.
```

FEATURES
 source
 1. 154280
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone.lib="RP11-63508"
 /clone.lib="RP11-11 Human Male BAC"
 1. 19698
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

```
misc_feature
19799..21284
/note="assembly_fragment"
21385..25070
/note="assembly_fragment"
25171..32681
/note="assembly_fragment"
32782..41653
/note="assembly_fragment"
41754..45817
/note="assembly_fragment"
45918..52511
/note="assembly_fragment"
52612..64034
/note="assembly_fragment"
64135..135344
/note="assembly_fragment"
135445..154280
/note="assembly_fragment"
clone_end:R7
vector_side:right"

BASE COUNT 43569 a 33643 c 33727 g 42441 t 900 others
ORIGIN
```

Query Match 87.0% Score 17.4; DB 2; Length 154280;
 Best Local Similarity 94.7%; Pred No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ggccttgctggccttgctt 20
 ||||||||||||||||
 Db 79877 GGCTTGCTGGCTTGCTT 79859

RESULT 8
 AC018956/c 155808 bp DNA HTG 16-MAR-2000
 LOCUS AC018956/c 155808 bp DNA HTG 16-MAR-2000
 DEFINITION Homo sapiens clone RP11-18F14, WORKING DRAFT SEQUENCE, 13 unordered

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 155808)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-18F14
 Unpublished
 2 (bases 1 to 155808)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Bouckhagter,B., Brown,A., Burkett,G., Castle,A.,
 Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearrellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karakas,A., Klein,J.,
 Landers,T., Lenoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 Meldrum,J., Menus,L., Morrow,J., Naylor,J., Norman,C.H.,
 O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K., Pierre,N.,
 Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
 Santos,R., Severy,P., Spencer,B., Stange-Rothman,N., Stojanovic,N.,
 Subramanian,A., Talamas,J., Testaye,S., Theodore,D., Tirelli,A.,
 Vasiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.D., Zimmer,A.
 and Zody,M.

TITLE
 JOURNAL
 COMMENT
 Submitted (24-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced g1:6721253.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 15331

Center clone name: 18_F_14

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 143382 bases at least Q40

Consensus quality: 149320 bases at least Q20

Consensus quality: 151893 bases at least Q20

Insert size: 153000; agarose-fp

Insert size: 154608; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 13 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```
1 1159: contig of 1159 bp in length
* 1160 1259: gap of 100 bp
* 1260 2778: contig of 1519 bp in length
* 2779 2878: gap of 100 bp
* 2879 6196: contig of 3318 bp in length
* 6197 6296: gap of 100 bp
* 6297 8665: contig of 2369 bp in length
* 8666 8765: gap of 100 bp
```

```
*      8766      12339: contig of 3574 bp in length
*      12340      12439: gap of 100 bp
*      12440      16387: contig of 3948 bp in length
*      16388      16487: gap of 100 bp
*      16488      25928: contig of 9441 bp in length
*      25929      26028: gap of 100 bp
*      26029      36430: contig of 10402 bp in length
*      36431      36530: gap of 100 bp
*      36531      48734: contig of 12204 bp in length
*      48735      48834: gap of 100 bp
*      48835      66567: contig of 17733 bp in length
*      66568      66667: gap of 100 bp
*      66668      94768: contig of 28101 bp in length
*      94769      94868: gap of 100 bp
*      94869      121753: contig of 26885 bp in length
*      121754      121853: gap of 100 bp
*      121854      155808: contig of 33955 bp in length.
FEATURES
Source
1. 153808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-18P14"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 1159
/note="assembly_fragment"
misc_feature
1260. 2778
/note="assembly_fragment"
misc_feature
2879. 6196
/note="assembly_fragment"
misc_feature
6297. 8665
/note="assembly_fragment"
misc_feature
8766. 12339
vector_side:right"
misc_feature
12440. 16387
/note="assembly_fragment"
misc_feature
12440. 16387
/note="assembly_fragment"
misc_feature
16488. 25928
vector_side:right"
misc_feature
26029. 36430
/note="assembly_fragment"
misc_feature
36531. 48734
/note="assembly_fragment"
misc_feature
48835. 66567
/note="assembly_fragment"
misc_feature
66668. 94768
/note="assembly_fragment"
misc_feature
94869. 121753
/note="assembly_fragment"
misc_feature
121854. 155808
/note="assembly_fragment"
BASE COUNT      46162 a      33931 c      33624 g      40888 t      1203 others
ORIGIN
Query Match      87.0%: Score 17.4: DB 2: Length 155808;
Best Local Similarity 94.7%: Pred. No. 1.4e+02;
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
KEYWORDS      HTG; HTGS_PHASE2; HTGS_CANCELED.
SOURCE
ORGANISM      Homo sapiens
REFERENCE
AUTHORS      Tracey/A.
TITLE      Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL      CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba34p4
----- Summary Statistics
Sequencing program: XGAP4: version 4.5
Sequencing vector: plasmid: 108752: 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 160070 bases at least Q40
Consensus quality: 160454 bases at least Q30
Consensus quality: 160732 bases at least Q20
Insert size: 161486; sum-of-contigs
Insert size: 144857; 5.9% error: agrose-fp
Quality coverage: 10.17x in Q20 bases; sum-of-contigs Quality
coverage: 11.34x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
1. 161486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-34P4"
/clone_lib="RPC1-11.1"
misc_feature
1. 161486
/note="assembly_fragment:02766"
misc_feature
161486. 25928
vector_side:right"
misc_feature
26029. 36430
/note="assembly_fragment"
misc_feature
36531. 48734
/note="assembly_fragment"
misc_feature
48835. 66567
/note="assembly_fragment"
misc_feature
66668. 94768
/note="assembly_fragment"
misc_feature
94869. 121753
/note="assembly_fragment"
BASE COUNT      54067 a      28905 c      28737 g      49777 t
ORIGIN
Query Match      87.0%: Score 17.4: DB 2: Length 161486;
Best Local Similarity 94.7%: Pred. No. 1.4e+02;
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 99gctgtgctggtgct 19
Db      29880 GGGCTTGCTGGCTTCT 29862

RESULT  9
LOCUS   AL590008      161486 bp      DNA      HTG      10-JUL-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-34P4, *** SEQUENCING IN
PROGRAMS *** , In ordered pieces.
ACCESSION AL590008
VERSION   AL590008.4      GI:13751585
```

```
RESULT  10
LOCUS   AC092332      166625 bp      DNA      HTG      03-JUL-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-18P14, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
ACCESSION AC092332
VERSION   AC092332.1      GI:14589521
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

REFERENCE 1 (bases 1 to 166625)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166625)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 413411
Center clone name: RPCI-11_18F14

Summary Statistics
Consensus quality: 150681 bases at least Q40
Consensus quality: 156633 bases at least Q30
Consensus quality: 157942 bases at least Q20
Estimated insert size: 147000; agarose-fp estimation
Estimated insert size: 165125; sum-of-contrigs estimation
Quality coverage: 10.18 in Q20 bases; agarose-fp estimation
Quality coverage: 9.06 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1032: contrig of 1032 bp in length
* 1033 1132: gap of unknown length
* 1133 2148: contrig of 1016 bp in length
* 2149 2248: gap of unknown length
* 2249 3774: contrig of 1526 bp in length
* 3775 3874: gap of unknown length
* 3875 4891: contrig of 1017 bp in length
* 4892 4991: gap of unknown length
* 4992 6048: contrig of 1057 bp in length
* 6049 6148: gap of unknown length
* 6149 7989: contrig of 1841 bp in length
* 7990 8089: gap of unknown length
* 8090 9310: contrig of 1221 bp in length
* 9311 9410: gap of unknown length
* 9411 11560: contrig of 2150 bp in length
* 11561 11660: gap of unknown length
* 11661 14602: contrig of 2942 bp in length
* 14603 14702: gap of unknown length
* 14703 16918: contrig of 2216 bp in length
* 16919 17018: gap of unknown length
* 17019 20585: contrig of 3567 bp in length
* 20586 20685: gap of unknown length
* 20686 29040: contrig of 8355 bp in length
* 29041 29140: gap of unknown length
* 29141 44499: contrig of 15359 bp in length
* 44500 44599: gap of unknown length
* 44600 67722: contrig of 23123 bp in length
* 67723 67822: gap of unknown length
* 67823 103449: contrig of 35627 bp in length
* 103450 103549: gap of unknown length
* 103550 166625: contrig of 63076 bp in length.
Location/Qualifiers
1. 166625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone-lib="RPCI human BAC library 11"
BASE COUNT 47807 a 35966 c 36005 g 45269 t 1578 others

ORIGIN
Query Match 87.0%; Score 17.4; DB 2: Length 166625;
Best Local Similarity 94.7%; Pred. No. 1.4e-02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gggctgctgggctgctg 19
|||||
Db 76728 GGGCTTGCTTGGGCTTCTT 76746

RESULT 11
AC024112
LOCUS
DEFINITION Mus musculus chromosome 11 clone RP23-277L11, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
AC024112
AC024112.14 GI:12025590
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM house mouse.
Mus musculus
REFERENCE 1 (bases 1 to 190460)
AUTHORS Metzker, M. L., Lewis, L. R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carllock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Bunay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louissege, H., Lozado, R. J., Martin, R.,
Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K., and Gibbs, R.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 190460)
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190460)
AUTHORS Worley, K. C.
JOURNAL Direct Submission
TITLE Submitted (24-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT -----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: MABY
Center clone name: RP23-277L11

Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-Primer Bodyipy: 96% of reads
Chemistry: Dye-terminator Big Dye: 4% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167285 bases at least Q40
Consensus quality: 178540 bases at least Q30
Consensus quality: 184007 bases at least Q20
Estimated insert size: 181421; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.5x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contrigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 15838: contig of 15838 bp in length
* 15839: gap of unknown length
* 15938: contig of 13305 bp in length
* 29243: gap of unknown length
* 29343: gap of unknown length
* 29344: gap of unknown length
* 44597: gap of unknown length
* 44697: contig of 11291 bp in length
* 55988: gap of unknown length
* 56087: contig of 13854 bp in length
* 69942: gap of unknown length
* 70042: contig of 11855 bp in length
* 81897: gap of unknown length
* 81997: gap of unknown length
* 92644: contig of 10648 bp in length
* 92645: gap of unknown length
* 92744: gap of unknown length
* 101175: contig of 8431 bp in length
* 101176: gap of unknown length
* 101275: contig of 10954 bp in length
* 112329: gap of unknown length
* 112330: contig of 7860 bp in length
* 112330: contig of 7860 bp in length
* 120189: gap of unknown length
* 120289: contig of 7705 bp in length
* 120290: gap of unknown length
* 127995: gap of unknown length
* 128095: contig of 6715 bp in length
* 134810: gap of unknown length
* 134810: gap of unknown length
* 134910: contig of 5959 bp in length
* 140868: gap of unknown length
* 140968: gap of unknown length
* 149273: contig of 8305 bp in length
* 149274: gap of unknown length
* 149374: contig of 5340 bp in length
* 154813: gap of unknown length
* 154813: gap of unknown length
* 159953: contig of 5140 bp in length
* 159954: gap of unknown length
* 160053: gap of unknown length
* 164797: contig of 4744 bp in length
* 164798: gap of unknown length
* 164897: gap of unknown length
* 164898: contig of 7151 bp in length
* 172049: gap of unknown length
* 172148: gap of unknown length
* 172149: contig of 2575 bp in length
* 174724: gap of unknown length
* 174824: gap of unknown length
* 177096: contig of 2273 bp in length
* 177097: gap of unknown length
* 177197: gap of unknown length
* 179215: contig of 2018 bp in length
* 179315: gap of unknown length
* 182115: contig of 2801 bp in length
* 182116: gap of unknown length
* 184215: gap of unknown length
* 184774: contig of 2559 bp in length
* 184775: gap of unknown length
* 184875: gap of unknown length
* 186397: contig of 1523 bp in length
* 186398: gap of unknown length
* 186498: gap of unknown length
* 187727: contig of 1230 bp in length
* 187828: gap of unknown length
* 189139: contig of 1312 bp in length
* 189140: gap of unknown length
* 189239: gap of unknown length
* 189240: contig of 1221 bp in length.
  
```

FEATURES

```

source
  1. 190460
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /chromosome="11"
  /clone="RP3-277L11"
BASE COUNT 58801 a 35584 c 35256 g 58200 t 2619 others
ORIGIN
  
```

```

Query Match      87.0%; Score 17.4; DB 2; Length 190460;
Best Local Similarity 94.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  
```

```

Oy 1 99gcttgcctgggcttgcct 19
Db 163879 GGGCTTGGCTTGGACTTGCCT 163897
  
```

```

RESULT 12
AC023948/C
LOCUS Homo sapiens chromosome 8 clone RP11-665B22 map 8, WORKING DRAFT
DEFINITION AC023948.3 GI:13184180
ACCESSION AC023948
VERSION AC023948.3
KEYWORDS HTG; PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
  
```

```

REFERENCE 1 (bases 1 to 196213)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-665B22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196213)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
  Boudgallier,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
  Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
  Deatellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
  Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
  Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
  Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
  Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
  Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
  Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
  McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
  Meneus,L., Mihov,T., Miranda,C., Miengo,V., Morrow,J., Naylor,J.,
  Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
  Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
  Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
  Severy,P., Spencer,B., Stage-Thomann,N., Stojanovic,N.,
  Subramanian,A., Talmas,J., Testaye,S., Theodore,J., Tirelli,A.,
  Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
  Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zimmer,A. and
  Zody,M.
  
```

```

TITLE Direct Submission
JOURNAL Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  
```

COMMENT

```

All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
  
```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
  
```

```

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
  
```

```

----- Project Information
Center project name: 16272
Center clone name: 669.B.22
  
```

```

----- Summary Statistics
Sequencing vector: M13; M77815; 59% of reads
Sequencing vector: Plasmid; n/a; 41% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 192380 bases at least Q40
Consensus quality: 194550 bases at least Q30
Consensus quality: 195428 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 195913; sum-of-contigs
Quality coverage: 7.4 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 b.
  
```

```

* NOTE: This is a 'working draft' sequence. It currently
  * consists of 4 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  
```

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
      1       20858: contig of 20858 bp in length
      *       20859 20958: gap of          100 bp
      *       118886: contig of 97928 bp in length
      *       118887 118986: gap of          100 bp
      *       118987 160470: contig of 41484 bp in length
      *       160471 160570: gap of          100 bp
      *       160571 196213: contig of 35643 bp in length.
FEATURES
SOURCE
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="8"
    /map="8"
    /clone="RP11-669B22"
    /clone_1bp="RPCR-11 Human Male BAC"
    1..20858
    /note="assembly_fragment
    clone_end:SP6
    vector_side:left"
misc_feature
    20959..118886
    /note="assembly_fragment"
    118987..160470
    /note="assembly_fragment"
    160571..196213
    /note="assembly_fragment
    clone_end:T7
    vector_side:right"
BASE COUNT   61413 a 37858 c 37706 g 58932 t     304 others
ORIGIN
Query Match      87.0% Score 17.4; DB 2; Length 196213;
Best Local Similarity 94.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
OY      2      gagctgacttgagcttgcctt 20
        ||||||
Db      26231 GTCTTGCTTGGGCTTGCTT 26213

RESULT 13
AL356310/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-354K1, *** SEQUENCING IN
PROGRESS ***, 17 unordered pieces.
ACCESSION AL356310
VERSION AL356310.4 GI:9797487
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 201126)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213833.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA35AKI
----- Summary Statistics
Assembly program: XGAP4; version 4.5
```

```

Sequencing vector: plasmid: 108753; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 191429 bases at least Q40
Consensus quality: 195113 bases at least Q30
Consensus quality: 197525 bases at least Q20
Insert size: 199526; sum-of-contigs
Insert size: 216699; 3.4% error; agarose-1p
Quality coverage: 3.65x in Q20 bases; sum-of-contigs Quality
coverage: 3.53x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
1
* 3012: contig of 3012 bp in length
* 3013 3112: gap of 100 bp
* 3113 19278: contig of 16166 bp in length
* 19279 19378: gap of 100 bp
* 19379 27470: contig of 8092 bp in length
* 27471 27570: gap of 100 bp
* 27571 38426: contig of 10856 bp in length
* 38427 38526: gap of 100 bp
* 38527 50499: contig of 11973 bp in length
* 50500 50599: gap of 100 bp
* 50600 53244: contig of 2645 bp in length
* 53245 53344: gap of 100 bp
* 53345 63794: contig of 10450 bp in length
* 63795 63894: gap of 100 bp
* 63895 67917: contig of 4023 bp in length
* 67918 68017: gap of 100 bp
* 68018 77251: contig of 9234 bp in length
* 77252 77351: gap of 100 bp
* 77352 113499: contig of 36148 bp in length
* 113500 113599: gap of 100 bp
* 113600 118617: contig of 5018 bp in length
* 118618 118717: gap of 100 bp
* 118718 143301: contig of 24584 bp in length
* 143302 143401: gap of 100 bp
* 143402 148150: contig of 4749 bp in length
* 148151 148250: gap of 100 bp
* 148251 155032: contig of 6782 bp in length
* 155033 155132: gap of 100 bp
* 155133 162200: contig of 7068 bp in length
* 162201 162300: gap of 100 bp
* 162301 182671: contig of 20371 bp in length
* 182672 182771: gap of 100 bp
* 182772 2011126: contig of 18355 bp in length.
Location/Qualifiers
1. 2011126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-354K1"
/clone_lib="RP11-354K1.2"
1. 3012
/note="assembly_fragment:00192
clone_end:77
vector_side:left"
3113. 19278
/note="assembly_fragment:00325
fragment_chain:1"
19379. 27470
/note="assembly_fragment:00479
fragment_chain:1"
27571. 38426
/note="assembly_fragment:01983
fragment_chain:1"
38527. 50499
/note="assembly_fragment:01045

```


Query Match	87.0%:	Score 17.4:	DB 2:	Length 205932:
Best Local Similarity	94.7%:	Pred. No. 1.3e+02:		
Matches 18:	Conservative 0:	Mismatches 1:	Indels 0:	Gaps 0
Oy	1	gggcttgcttgggcttgcct 19		
Db	101373	GGGATGTGCTTGGGCTTGCTT 101391		
RESULT 15				
AC073742				
LOCUS				
DEFINITION	AC073742 215653 bp	HTG	18-JUL-2000	
	Mus musculus clone RP23-27D5,	WORKING DRAFT SEQUENCE,	14 ordered	
	pieces.			
ACCESSION	AC073742			
VERSION	AC073742.2	GI:9256780		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 215653)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Sequencing of Mouse			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 215653)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Direct Submission			
JOURNAL	Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint			
	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
	On Jul 18, 2000 this sequence version replaced g1:8810359.			
COMMENT				

```

Project Information
Center Project Name: 1747317
Center clone name: RPEC1-23_27DS
-----
Summary Statistics
Consensus quality: 203285 bases at least Q40
Consensus quality: 211686 bases at least Q30
Consensus quality: 213256 bases at least Q20
Estimated insert size: 233000; agarose-1p estimation
Estimated insert size: 215053; sum-of-contigs estimation
Quality coverage: 7.23 in Q20 bases; agarose-1p estimation
Quality coverage: 7.84 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
  This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
12377 12376: contig of 12376 bp in length
12477 12476: gap of unknown length
55546: contig of 43070 bp in length
55647 55646: gap of unknown length
53647 71884: contig of 16238 bp in length
71885 71884: gap of unknown length
71985 86752: contig of 14768 bp in length
86753 86852: gap of unknown length
86853 96739: contig of 9887 bp in length
96740 96839: gap of unknown length
96840 105439: contig of 8600 bp in length
105440 105539: gap of unknown length
105540 110329: contig of 4790 bp in length
110330 110429: gap of unknown length
110430 115111: contig of 4682 bp in length
115112 115211: gap of unknown length

```

```

Query Match          87.0%:  Score 17.4:  DB 2:  Length 215653;
Best Local Similarity 94.7%:  Pred. No. 1.3e+02;
Matches 18:  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0.

QY      1  ggcgtctgcttggcgttgc 19
      ||| ||||| ||||| |||||
Db 55099 GGGATTGCTGGGGCTTGCT 55117

Search completed: December 8, 2001, 10:52:16
Job time: 5001 sec

```

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:50:54 ; Search time 1723.37 Seconds

(without alignments)
134.017 Million cell updates/sec

Title: US-09-508-147-10

Perfect score: 14

Sequence: 1 cttcgttgacacag 14

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	14	100.0	961	2	AC063109	AC063109 Giardia i
C 2	14	100.0	968	2	AC052567	AC052567 Giardia i
C 3	14	100.0	1588	5	AF092363	AF092363 Cynopoei
C 4	14	100.0	2615	8	YSCAARI	L12542 Yeast alani
C 5	14	100.0	39101	1	AF069392	AF069392 Vibrio pa
C 6	14	100.0	44113	8	YSCHE179	U00062 Saccharomyc
7	14	100.0	101500	9	AL365505	AL365505 Human DNA
8	14	100.0	123385	9	AL359711	AL359711 Human DNA
9	14	100.0	128598	1	D90912	D90912 Synechocyst
C 10	14	100.0	150447	2	AL136222	AL136222 Homo sapi
C 11	14	100.0	158469	2	AC021608	AC021608 Homo sapi
C 12	14	100.0	163526	8	AF002522	AF002522 Oryza sat
C 13	14	100.0	164879	9	AC007876	AC007876 Homo sapi
C 14	14	100.0	166955	2	AC080161	AC080161 Homo sapi
C 15	14	100.0	167029	8	AF002540	AF002540 Oryza sat
C 16	14	100.0	177014	2	AC022269	AC022269 Homo sapi
C 17	14	100.0	179589	2	AC018860	AC018860 Homo sapi
C 18	14	100.0	182510	9	AL356278	AL356278 Human DNA
C 19	14	100.0	186230	2	AC024613	AC024613 Homo sapi
C 20	14	100.0	192208	2	AC074297	AC074297 Homo sapi
C 21	14	100.0	193128	2	AL138744	AL138744 Homo sapi
C 22	14	100.0	203504	2	AC068162	AC068162 Homo sapi
C 23	14	100.0	216456	2	AC026227	AC026227 Homo sapi
C 24	14	100.0	221284	2	AC053502	AC053502 Homo sapi
C 25	14	100.0	224574	2	AC069153	AC069153 Homo sapi
C 26	14	100.0	224574	2	AC069153	AC069153 Homo sapi
C 27	14	100.0	302	11	G23076	G23076 human STS W
C 28	13	92.9	342	14	AF238137	AF238137 Hepatitis
C 29	13	92.9	342	14	AF238157	AF238157 Hepatitis
C 30	13	92.9	342	14	AF238158	AF238158 Hepatitis
C 31	13	92.9	342	14	AF238213	AF238213 Hepatitis
C 32	13	92.9	379	11	G30767	G30767 SWSS3180 ET
C 33	13	92.9	336	6	AF108940	AF108940 Mercenari
C 34	13	92.9	504	8	AF306534	AF306534 Coccidioid
C 35	13	92.9	565	5	AF146675	AF146675 Chlorella
C 36	13	92.9	805	9	HSANKME38	U50129 Homo sapien
C 37	13	92.9	906	8	SA4036	SA4036 mstr-1-multi
C 38	13	92.9	1018	9	AY034471	AY034471 Homo sapi
C 39	13	92.9	1023	9	HS8802540	AL161963 Homo sapi
C 40	13	92.9	1054	10	AB049227	AB049227 Rattus nor
C 41	13	92.9	1100	8	AF239989	AF239989 Prunus pe
C 42	13	92.9	1194	8	AF385741	AF385741 Arabidops
C 43	13	92.9	1315	9	AF151812	AF151812 Homo sapi
C 44	13	92.9	1342	9	BC009765	BC009765 Homo sapi
C 45	13	92.9	1446	6	A52418	A52418 Sequence 1

ALIGNMENTS

RESULT	1	HTG	21-APR-2000
AC063109/c	AC063109	961 bp	DNA
LOCUS	Giardia intestinalis clone NJ3102	strain WB-C6,	LOW-PASS SEQUENCE
DEFINITION	AC063109	961 bp	DNA
SAMPLING	AC063109	961 bp	DNA
ACCESSION	AC063109.1	GI:7633768	
VERSION	HTG: HTGS-PHASE0.		
KEYWORDS	HTG: HTGS-PHASE0.		
SOURCE	Giardia intestinalis		
ORGANISM	Giardia intestinalis		
REFERENCE	Eukaryota; Diplomonadida; Hexamitidae; Giardia		
AUTHORS	1 (bases 1 to 961)		
TITLE	Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.Q., Kim, U.,		
JOURNAL	Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.		
REFERENCE	Giardia: a model for ancient eukaryotic genome analysis		
AUTHORS	2 (bases 1 to 961)		
TITLE	Unpublished		
REFERENCE	Nixon, J., Morrison, H.G., McArthur, A.G., Eakin, N.Q., Kim, U.,		
AUTHORS	Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.		
TITLE	Direct Submission		

JOURNAL
Submitted (21-APR-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT
* NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 961: contig of 961 bp in length.
FEATURES
source
1..961
Location/Qualifiers
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="NJ3102"
BASE COUNT
273 a 236 c 231 g 221 t
ORIGIN

Query Match 100.0%; Score 14; DB 2; Length 961;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttaacagg 14
|||||
Db 373 cttgcttaacagg 360

RESULT 2
AC052567/c
LOCUS
DEFINITION
AC052567 968 bp DNA HTG 15-APR-2000
AC052567
AC052567.1 GI:7575887
HTG: HTGS PHASE0
KEYWORDS
SOURCE
ORANISM
Giardia intestinalis
REFERENCE
AUTHORS
1 (bases 1 to 968)
Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O., Kim, U., Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.
Giardia: a model for ancient eukaryotic genome analysis
Unpublished
2 (bases 1 to 968)
Kim, U., Morrison, H.G., McArthur, A.G., Nixon, J., Eakin, N.O., Crocker, M.C., Hinkle, G., Holder, M.E. and Sogin, M.L.
Direct Submission
Submitted (15-APR-2000) Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT
* NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 968: contig of 968 bp in length.
FEATURES
source
1..968
Location/Qualifiers
/organism="Giardia intestinalis"
/strain="WB-C6"

/db_xref="taxon:5741"
/clone="KJ3702"
BASE COUNT 272 a 228 c 254 g 211 t 3 others
ORIGIN

Query Match 100.0%; Score 14; DB 2; Length 968;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttaacagg 14
|||||
Db 652 cttgcttaacagg 639

RESULT 3
AF092363
LOCUS
DEFINITION
AF092363 1588 bp DNA VRT 19-OCT-1999
Cynopoeclius melanoaenia trna-ile gene, partial sequence; trna-gln and trna-met genes, complete sequence; nadh dehydrogenase subunit II gene, complete cds; trna-trp, trna-ala, trna-asn, trna-cys, and trna-tyr genes, complete sequence; and cytochrome c oxidase subunit I gene, partial cds; mitochondrial genes for mitochondrial products.
AF092363
AF092363.1 GI:6062671
KEYWORDS
SOURCE
ORGANISM
Cynopoeclius melanoaenia.
Mitochondrion Cynopoeclius melanoaenia
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorph; Atherinomorpha; Cyprinodontiformes; Aplocheilidae; Rivulinae; Cynopoeclius.
1 (bases 1 to 1588)
Hrbek, T. and Larson, A.
The evolution of diapause in the killifish family Rivulidae (Atherinomorpha, Cyprinodontiformes): A molecular phylogenetic and biogeographic perspective
Evolution 53 (4), 1200-1216 (1999)
2 (bases 1 to 1588)
Hrbek, T. and Larson, A.
Direct Submission
Submitted (17-SEP-1998) Dept. of Biology, Washington University, Box 1137, St. Louis, MO 63130, USA
1..1588
Location/Qualifiers
/organism="Cynopoeclius melanoaenia"
/organella="mitochondrion"
/db_xref="taxon:105041"
<1..4
/product="trna-ile"
complement(4..74)
/product="trna-gln"
74..142
/product="trna-met"
143..1189
/transl_table=2
/codon_start=1
/product="nadh dehydrogenase subunit II"
/protein_id="AA02924.1"
/db_xref="GI:6062746"
/translation="MPCVFIIILIMMGITMTLSSHWLAWTGLEINTFAIPLM TODKHPAMAAALRYFLVOSTAATLTLFAATSNAMSGOEIOOMTALPLTMAIIL IIMKGLAPLHSMPLREVLOGINLTGILSTWOKIAPLAILIOTSYSPSAIVAGILS ILIGWGGLNQTOKITLAIASSITIHGMILIIHYAPLAPMSIATYMTATAPYMF SYLQTKMNSLFTSWNSPLDFSTVFLSLDGLGPPFSGVPRKMLILQEMAOALMP LATMAALFSLISLYYLRLSYFMAALMPGNLPATLSMPCRPPLTLTALIVFSLC LPLIPSFNFFIY"
1188..1259
/product="trna-trp"
complement(1260..1328)
/product="trna-ala"
complement(1329..1400)

```

rep_origin      /product="tRNA-Asn"
                1405. .1436
                /note="light strand"
tRNA            complement(1437. .1502)
                /product="tRNA-Cys"
                complement(1505. .1568)
tRNA            /product="tRNA-Tyr"
                1572. .>1588
CDS             /codon_start=1
                /transl_table=2
                /product="cytochrome c oxidase subunit I"
                /protein_id="AA02925.1"
                /db_xref="GI:6062747"
                /translation="MTVAR"
BASE COUNT     446 a 431 c 217 g 494 t
ORIGIN
Query Match    100.0%; Score 14; DB 5; Length 1588;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
Db 224 CTTGCTTGACAGG 237

RESULT 4
YSCAPI/c       YSCAPI 2615 bp DNA PLN 04-AUG-1993
LOCUS          Yeast alanine/arginine aminopeptidase (AAP1) gene sequence.
DEFINITION     L12542
ACCESSION      L12542.1 GI:295571
VERSION        AAP1 gene; alanine/arginine aminopeptidase; aminopeptidase.
KEYWORDS       Saccharomyces cerevisiae (strain ) DNA.
SOURCE         Saccharomyces cerevisiae
ORGANISM       Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE      1 (bases 1 to 2615)
AUTHORS       Caprioglio,D.R., Padilla,C. and Werner-Washburne,M.
TITLE         Isolation and characterization of AAP1: A gene encoding an
               alanine/arginine aminopeptidase in yeast
               J. Biol. Chem. 268, 14310-14315 (1993)
JOURNAL        93300827
MEDLINE
FEATURES       location/Qualifiers
               source
               1..2615
               /organism="Saccharomyces cerevisiae"
               /strain="DS10"
               /db_xref="taxon:4932"
BASE COUNT     796 a 457 c 543 g 819 t
ORIGIN
Query Match    100.0%; Score 14; DB 8; Length 2615;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
Db 189 CTTGCTTGACAGG 176

RESULT 5
LOCUS          AF069392 39101 bp DNA BCT 24-MAR-2000
DEFINITION     Vibrio parahaemolyticus polar flagellar and chemotaxis genes,
               complete cds; and unknown genes.
ACCESSION      AF069392 U12816 U36917
VERSION        AF069392.3 GI:7327274
KEYWORDS       Vibrio parahaemolyticus.
SOURCE         Vibrio parahaemolyticus
ORGANISM       Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

```

```

REFERENCE      1 (bases 1333 to 6861)
AUTHORS       McCarter,L.L.
TITLE         Genetic and molecular characterization of the polar flagellum of
               Vibrio parahaemolyticus
JOURNAL        J. Bacteriol. 177 (6), 1595-1609 (1995)
MEDLINE        95189739
PUBMED         7883718
REFERENCE      2 (bases 6862 to 9675)
AUTHORS       Stewart,B.J. and McCarter,L.L.
TITLE         Vibrio parahaemolyticus FlaA, a homologue of FlhS, is required for
               production of a flagellin
JOURNAL        Mol. Microbiol. 20 (1), 137-149 (1996)
MEDLINE        97014377
PUBMED         8861212
REFERENCE      3 (bases 1 to 39101)
AUTHORS       Boles,B.R. and McCarter,L.L.
TITLE         Insertional inactivation of genes encoding components of the
               sodium-type flagellar motor and switch of vibrio parahaemolyticus
JOURNAL        J. Bacteriol. 182 (4), 1035-1045 (2000)
MEDLINE        20115548
PUBMED         10648530
REFERENCE      4 (bases 1 to 39101)
AUTHORS       Jaques,S., Kim,Y.K. and McCarter,L.L.
TITLE         Components of the polar flagellar switch complex and assembly
               apparatus
JOURNAL        Unpublished
REFERENCE      5 (bases 1 to 9674)
AUTHORS       McCarter,L.L.
TITLE         Direct Submission
               Submitted (01-AUG-1994) Immunology, The Scripps Research Institute,
               10666 North Torrey Pines Rd., La Jolla, CA 92037, USA
JOURNAL        6 (bases 12010 to 27963)
AUTHORS       Kim,Y.K. and McCarter,L.L.
TITLE         Direct Submission
               Submitted (30-MAY-1998) Microbiology, University of Iowa, Bowen
               Science Building, Iowa City, IA 52242, USA
JOURNAL        7 (bases 1 to 9674)
AUTHORS       McCarter,L.L.
TITLE         Direct Submission
               Submitted (30-JUL-1998) Microbiology, University of Iowa, Bowen
               Science Building, Iowa City, IA 52242, USA
JOURNAL        8 (bases 1 to 39101)
AUTHORS       Kim,Y.K. and McCarter,L.L.
TITLE         Direct Submission
               Submitted (09-FEB-2000) Microbiology, University of Iowa, Bowen
               Science Building, Iowa City, IA 52242, USA
JOURNAL        9 (bases 1 to 39101)
AUTHORS       Kim,Y.K. and McCarter,L.L.
TITLE         Direct Submission
               Submitted (24-MAR-2000) Microbiology, University of Iowa, Bowen
               Science Building, Iowa City, IA 52242, USA
REMARK         Sequence update by submitter
COMMENT        On Mar 24, 2000 this sequence version replaced gi:6806922
               gi:6952810.
FEATURES       location/Qualifiers
               source
               1..39101
               /organism="Vibrio parahaemolyticus"
               /strain="B82"
               /db_xref="taxon:670"
               371..1504
               /gene="flaF"
               371..1504
               /gene="flaF"
               /codon_start=1
               /transl_table=11
               /product="polar flagellin"
               /protein_id="AAC27808.1"
               /db_xref="GI:3353298"
               /translation="MAITVNTNVAALVAORHLTSATMDLNSMERLSSGKRINSKAD
               AAGIQISNRLOSOMGLDVAARNANDGISIQTAEGANNEVTNIMQRRDLISQANG
               SNSQERRPALDEVTALNDELRIIAETTSFGKRLNLTGTFKSSPQICAGSEAVQIE

```


TITLE Willis, A., Wilson, R., Wohlman, P. and Waterston, R.
Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII
JOURNAL Science 265 (5181), 2077-2082 (1994)
MEDLINE 94378003
PUBMED 8091229
REFERENCE 2 (bases 1 to 44113)
AUTHORS Du, Z.
TITLE The sequence of *S. cerevisiae* cosmid 8179
JOURNAL unpublished (1994)
REFERENCE 3 (bases 1 to 44113)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1994)
REFERENCE 4 (bases 1 to 44113)
AUTHORS Jia, Y. and Cherry, J.M.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1997) Department of Genetics, Stanford
University, Saccharomyces Genome Database, Stanford, CA 94305-5120,
USA

COMMENT
Curated by:
Saccharomyces Genome Database
URL: <http://genome-www.stanford.edu/>
e-mail: yeast-curator@genome.stanford.edu

Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: mj@sequencer.wustl.edu

NEIGHBORING COSMID INFORMATION:
1. 44113
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/map="VIII"
/strain="S288C (AB972)"
complement(539..1993)
/gene="Slr2"
complement(539..1993)
/gene="Slr2"
/note="YHR030C; Other names MPK1, SLK2, BYC2."
/codon_start=1
/product="Putative Ser/Thr protein kinases"
/protein_id="AAB68912.1"
/db_xref="GI:488177"
translation="MADKIERHTEFVNQDSVDKRFOLIKEIGHGAVGYCSARFAE
AAEDTVAIKKVTNVESEKTLCKRSRLKLRHGRGHKNITCLYMDIVIVYPOSSIN
GLIYELMECDMDIISGQPLIDAHQSTYQLGCKLTHSDVLRHDKPEENL
VNADQLKICDGLARGTSENVENSQFLTEVATRWTRAPBIMLSYGYTKADVWS
AGCIIAEFLGKPIFGKDYVNLQMLQVLTGPEDETLRRIGSKNVDYIHLQELIP
KYPEVNLNPNANSQALIDLEQMLAFDPKRTIVDEALHPYLSIWNHDADEPVCSEK
EFSFESVNDMEKQVIOEODFLFVROPLEBROLOLOOQOQOQOQOQOQOQOQOQ
SDVDGNAAASENPROMATSNVAPQOESFGIHSQNLPRHNDAPRRPQESMMEMR
PATGNTADIPQNDGTLTDLLEKLEFGLDKRY"
complement(2448..4619)
/gene="YHR031C"
complement(2448..4619)
/gene="YHR031C"
/note="Similar to Pif1p (*S. cerevisiae*), mitochondrial DNA
repair and recombination protein."
/codon_start=1
/evidence="not_experimental"
/product="Yhr031Cp"
/protein_id="AAB68913.1"

FEATURES
source
This sequence includes nucleotides 1-43838 of cosmid 8179 plus 75
bp of cosmid 9181 (cosmid adjacent to the right end of 8179). This
sequence overlaps with cosmid YSCH8025 on the right by 200 bp. The
adjacent cosmid to the left is YSCH8082.

gene
/db_xref="GI:488178"
translation="MFRSHASGNKKOMKSRSSNGSTPAPASGSHAYRQOTLSPEFG
/translating="MFRSHASGNKKOMKSRSSNGSTPAPASGSHAYRQOTLSPEFG
CGKSAASKNSTTIDLESQDEGRNRTAPRPLIINNSSGSLQSGSDGDDPD
AEFKIVDPLNLSKSKSRSLMSSTLAKTASASTQKTHPDEDFLREYTSKSN
SRQSTFTINTEDSSKLTSDSERPARSRPMEFQGLTPVKKIPLRLKTSYNS
DSMNRHSASPVLTMEQERYVNLIVKRTNVEFTYSGAGSKVLLQTIHQLSLYG
KESIAITASTGLAAVTIGSTLHKWSGIGIKTIDQLVKRKIOSQKDLAAMRYKVL
IIDESVMDGLDLKLEQIARIRKNDPFGIOLVTFGDFOLPPVAKKDEHNVK
IDESVMDGLDLKLEQIARIRKNDPFGIOLVTFGDFOLPPVAKKDEHNVK
CPSESMKRCIOKTLITLRFVRODNKTLIDILNAIRGELVADIAKTRINRQIDVA
DGIAPELTATRRREVELSNVKKLTSPDIDEFRAVDNAPERYQAIIDSLMEVVA
LKEDQVMLNKNKPDVELVNSGLKVLTFVESLVKKKEITKIVDDEVMDKRLSVR
VIGNPLKESKEFRODLNARPLARLEKILINAVAKSPKREPFVWVYKWKYI
HELVWPEREPIDIPRENVGLERTQIPLMICALSIHKAQGTOTIGRLKVDLRIFEAG
YVVALSRVATDQLVLFNPDGKITRNERVDFEYRLFTLK"
4993..6738
/gene="YHR032w"
4993..6738
/gene="YHR032w"
/db_xref="GI:488176"
translation="MSKQFSTTNDRRSIIYSTSVGKAGLFTPADYIPQSEENLIE
GEOGSEEPSYTGNDDETEREGEYHSLDANNRSTLQOAMQOQSDSHDKRLIDE
ERDLIDNKLLOHONGGDDIESHGGAIGDEERPAEIAINTESAIESGOKISTP
FKREOVITMNAPLIIPFIIQNSLSIASISVSILGTEGIGVITGSMPTANICLAA
IOGCTCIDTICAOATYGAKNHVLGVIVORCAVITTLAPLMRYWFWWSKILALMI
PERELCALAANLRLVATFAGVPGFIFECGKRRLQCGIFHASTIVLFCVADPLNANLY
LLVWMDKIGIGILGAPLSVINYMLTGLLIVMTYTHKSRPLCMNGIILPKQAFK
NMKMINLAIPGVWVAEAFLEGLVITFASHLIGDALGAOSIVATIASLAVYVFSI
SVSTSTRVANFISGLVDSQMITCVSLISFVSSMMMPFYICRYKREOIASLFTESA
VYKMWVCPPLIAPLQFDPALFNASTAGLCRQGRKKNWVQPSKILLPRCAHGICVS
IPVSGCRRLWGIN"
7197..8468
/gene="YHR033w"
7197..8468
/gene="YHR033w"
/note="highly similar to Prolp (Gamma-glutamyl kinase)"
/codon_start=1
/evidence="not_experimental"
/product="Yhr033wp"
/protein_id="AAB68910.1"
/db_xref="GI:488175"
translation="MTKAVTVYIKGSSLYDESTKEPKLSIMLYETVYNLRMGH
KLIIVSSGGLAVGDLALNIPRKPOLSEVOAIAAGQRLIARMMMLFSQGEQTAQI
LITRNDILRMQYNNARNITINELAMGVIPVENDTSLISEIERGNDTLTSAITAI
VGADFLFLTVDDCLYTNDPNRPDIIVPELSEGLPGVNTSSGSGSEYGTGMR
TKLIAADLASNAGIETIYMKSDRPEYVKIYDIOHHRPFRHNGTNGOQFLDQDT
ELBQIRKIDYPMHRTFLANDKKHKLKNSPMTLHGLITKGAIIIDENSYDKLSDMA
SLTPNAVILEVDNHEHLCVDLKTGORLPNSELDSKFIQSGVGVRSNTSLDLAKIK
GLPSEKHIDVLGYSEVVAHRENTAFPPQ"
complement(8614..9648)
/gene="YHR034C"
complement(8614..9648)
/gene="YHR034C"
/codon_start=1
/evidence="not_experimental"
/product="Yhr034cp"
/protein_id="AAB68914.1"
/db_xref="GI:488179"
translation="MADFLAPRQHRHNEKYYVADAAGDSVSKIEPIPAFKTKL
LSANGPEKLOGRKVFYNCHSPVLYPREVFNARIYPLIIQNMETPIITSCYRMD
HDKGQECYVWDCINSDCSRWICDDIOLREILVEMCLESEIRSDSVVLCKDRALPK
MKRGAELPALEVINDELADQYKAKMHIIEEGADPDSILGRNDDDDNNDDPGT
LPPLPTEINKISGAKIEIDIKNEIARMLKQAPAPAPHPHQQDEPEYEVKMKRFG
AAYKRLILIEKKAKPSKDRSPSYNFAENILYINGKLSIPLPRDIYVNAADIKIFHI
RKERTLYIYI"
9868..11760
/gene="YHR035w"
9868..11760
/gene="YHR035w"
/note="Weak similarity to Sec23p (*S. cerevisiae*)."

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr20>

RP11-382A12 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-382A12 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-332M is at 101401 in this sequence. The true right end of clone RP11-132F21 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with a alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Location/Qualifiers

1. 101500
 - /organism="Homo sapiens"
 - /db_xref="taxon:9606"
 - /chromosome="20"
 - /clone="RP11-382A12"
 - /clone_11b="RPCI-11.2"
2. 40
 - /note="Aluub repeat: matches 107. .149 of consensus"
41. .335
 - /note="AluY repeat: matches 1. .238 of consensus"
336. .520
 - /note="Aluub repeat: matches 149. .304 of consensus"
544. .841
 - /note="Alusx repeat: matches 1. .298 of consensus"
844. .1029
 - /note="MIR repeat: matches 49. .257 of consensus"
1069. .1185
 - /note="FLAM.C repeat: matches 1. .118 of consensus"
- complement(join(<1314. .1429,2557. .2648,6818. .6926,7045. .7145,9460. .9615,17278. .17348,20856. .20971,25125. .25285,31135. .31207,36834. .36900,41532. .41669))
 - /gene="dJ132F21.3"
 - /note="match: cDNAs: Em:AF228421 Em:U15635 Em:AL050267 Em:AK024353 Em:AF147427
 - match: ESTs: Em:BB888325 Em:AW058545 Em:AA3811682 Em:AT734982 Em:AI744621 Em:AW504769 Em:R25510 Em:AA496236 Em:AA381555 Em:M66047 Em:AA542149 Em:BB818948 Em:BE0061717 Em:BB818922 Em:W20038 Em:AA581478 Em:AA381497 Em:AA381599 Em:BE006144 Em:AA478751 Em:BE843840 Em:AI341137 Em:BE671955 Em:HA7862 Em:AI675309 Em:AW277012"
 - /evidence="not_experimental"
 - /product="ba382A12.1 (72.1 kDa protein (DKFZP564A032, SBB188) similar to mouse IFN-gamma induced MG11)"
 - complement(1314. .41669)
 - /gene="dJ132F21.3"
 - complement(join(<1314. .1429,2557. .2648,6818. .6926,7045. .7145,9460. .9615,17278. .17348,20856. .20971,25125. .25285,31135. .31207,36834. .36900,41532. .41739))
 - /gene="dJ132F21.3"
 - /note="continues in dJ132F21 (AL079335)"
 - match: cDNAs: Em:AF228421 Em:U15635 Em:AL050267 Em:AK024353 Em:AF147427
 - match: ESTs: Em:BB888325 Em:AW058545 Em:AA3811682 Em:AT734982 Em:AI744621 Em:AW504769 Em:R25510 Em:AA496236

Em:AA381555 Em:M86047 Em:AA542149 Em:BE818948 Em:BE006171
 Em:BE818922 Em:W20038 Em:AA381478 Em:AA381497 Em:AA381593
 Em:BE006144 Em:AA478751 Em:BE843840 Em:AI341137
 Em:BE671955 Em:H47862 Em:AI675309 Em:AM277012
 match: proteins: Tr:Q9Y3Z3 Sw:Q60710 Sw:Q09374 Tr:Q9VWY4
 Tr:Q26251 Tr:Q28849 Tr:P95966 Tr:Q58554 Tr:Q9V1T1
 Tr:Q9X1N3 Tr:Q59446 Tr:Q67745"
 /codon_start=1
 /evidence=not experimental
 /product="ba382a12.1 (72.1 kDa protein (DKFZ564A032,
 SBI188) similar to mouse IFN-gamma induced MG11)"
 /protein_id="CAC19806.1"
 /db_xref="GI:12043445"
 /translation="MORADSEQPSKRRCDSPPTSPNTSPAADWSPEGLHPDYKT
 WSPQVCSFLRRGGRPEEPLKINRENTGALLCLCPSEPRENTGVSSGRKLLS
 YIQRLVQIHDVDMKYINDPIHGHIEHLVITITTPQRLRYTKQLGGGTYRFGA
 SHNRRESLGAGVYLAGCLVHALGKQPELOISERVLCVLAGLHDHIGFSPMF
 GFPIFLAREVYKTHGQSVMMFEHLINSNGIKPYMEQGLIPEDICFIKQIVGPL
 ESPVEDSLMPYKGRPNKSFLEYISNKRNGIDVDKMFADCHHGLGIONFDYKRF
 IKFAVCEVDNELRICARDEKVENLYDMETHNSLHRRAYQHKVGNIIIDMTDAFLK
 ADDYIEITGAGCKKRYISAIIDMEATYTKLT"
 1604..1717
 repeat_region /note="MIR repeat: matches 116..238 of consensus"
 /note="2849..2934
 repeat_region /note="FLAM_A repeat: matches 43..142 of consensus"
 2992..3049
 repeat_region /note="Alu repeat: matches 243..300 of consensus"
 3083..3392
 repeat_region /note="Alu repeat: matches 1..309 of consensus"
 /note="complement(3140..3337)
 misc_feature /gene="d1132F21.3"
 /note="match: GSS: Em:AQ668484"
 complement(3157..3379)
 misc_feature /gene="d0132F21.3"
 /note="match: GSS: Em:AQ601622"
 complement(3165..3343)
 misc_feature /gene="d1132F21.3"
 /note="match: GSS: Em:AQ631997"
 3418..3716
 repeat_region /note="AluX repeat: matches 1..300 of consensus"
 3545..3684
 misc_feature /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 3719..3831
 repeat_region /note="MIR repeat: matches 148..254 of consensus"
 3853..4149
 repeat_region /note="AluX repeat: matches 1..297 of consensus"
 4241..4441
 repeat_region /note="Tiger2a repeat: matches 1..219 of consensus"
 4442..4755
 repeat_region /note="AluX repeat: matches 1..305 of consensus"
 4756..4824
 repeat_region /note="Tiger2a repeat: matches 219..293 of consensus"
 4825..5125
 repeat_region /note="AluY repeat: matches 5..300 of consensus"
 5126..5157
 repeat_region /note="Tiger2a repeat: matches 293..326 of consensus"
 5158..5449
 repeat_region /note="AluY repeat: matches 1..298 of consensus"
 5450..5550
 repeat_region /note="Tiger2a repeat: matches 326..435 of consensus"
 5551..5857
 repeat_region /note="AluX repeat: matches 1..307 of consensus"
 5858..5927
 repeat_region /note="Tiger2a repeat: matches 435..434 of consensus"
 5974..6267
 repeat_region /note="AluY repeat: matches 1..292 of consensus"
 6274..6445
 repeat_region /note="MIR repeat: matches 29..230 of consensus"
 6447..6573
 repeat_region /note="FLAM_A repeat: matches 1..127 of consensus"
 7468..7763
 repeat_region /note="AluX repeat: matches 1..297 of consensus"

repeat_region 8116..8346
 /note="AluY repeat: matches 1..294 of consensus"
 repeat_region 8433..8560
 /note="FLAM_C repeat: matches 1..128 of consensus"
 repeat_region 8613..8927
 /note="AluX repeat: matches 1..312 of consensus"
 repeat_region 8951..9240
 /note="AluY repeat: matches 2..291 of consensus"
 repeat_region 10228..10491
 /note="AluY repeat: matches 21..282 of consensus"
 repeat_region 10580..10877
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region 10896..11206
 /note="AluY repeat: matches 1..309 of consensus"
 repeat_region 11209..11508
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region 11681..11783
 /note="FLAM_A repeat: matches 1..123 of consensus"
 11784..11786
 /note="single clone region. Assembly confirmed by
 restriction digest data."
 repeat_region 11787..12081
 /note="AluX repeat: matches 4..307 of consensus"
 repeat_region 12090..12499
 /note="LIMC4 repeat: matches 7373..7773 of consensus"
 repeat_region 12500..12774
 /note="AluY repeat: matches 1..282 of consensus"
 repeat_region 12775..12785
 /note="LIMC4 repeat: matches 7773..7782 of consensus"
 repeat_region 12786..13044
 /note="AluY repeat: matches 1..293 of consensus"
 Query Match 100.0%; Score 14; DB 9; Length 101500;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ctgactgaacag 14
 Db 93551 CTTCCTTGACACG 93564
 RESULT 8
 AL359711
 LOCUS Human DNA sequence from clone RP11-425D10 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL359711
 VERSION AL359711.18 GI:13234940
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 123385)
 AUTHORS Tracey/A.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT
 On Mar 5, 2001 this sequence version replaced gi:12956944.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> Rpl1-425D10 is from the library RPlC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RPl1-425D10 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone Rpl1-425D10 is at 1 in this sequence. The true left end of clone RPl5-919F19 is at 123286 in this sequence. The true right end of clone Rpl1-35204 is at 45823 in this sequence.

FEATURES

source

1. 123385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Rpl1-425D10"
/clone_lib="RPlC1-11.2"
733. 1288
repeat_region
/note="L2 repeat: matches 1304. 1904 of consensus"
1289. 1586
repeat_region
/note="AluX repeat: matches 1. 297 of consensus"
1587. 1988
repeat_region
/note="L2 repeat: matches 1904. 2292 of consensus"
2405. 3144
repeat_region
/note="LIME2 repeat: matches 5348. 6155 of consensus"
3359. 3741
repeat_region
/note="MLT1B repeat: matches 4. 390 of consensus"
5536. 5828
repeat_region
/note="AluSc repeat: matches 1. 291 of consensus"
5830. 6144
repeat_region
/note="AluX repeat: matches 1. 307 of consensus"
7864. 8278
repeat_region
/note="LTR40a repeat: matches 1. 504 of consensus"
9928. 10104
repeat_region
/note="MER91B repeat: matches 1. 183 of consensus"
10243. 10388
repeat_region
/note="L2 repeat: matches 2349. 2498 of consensus"
10660. 10825
repeat_region
/note="L1MB3 repeat: matches 5983. 6143 of consensus"
11379. 11673
repeat_region
/note="AluSg repeat: matches 1. 295 of consensus"
11704. 12009
repeat_region
/note="AluX repeat: matches 1. 307 of consensus"
12687. 13467
misc_feature
/note="Cpg island"
/evidence="not_experimental"
13696. 13798
repeat_region
/note="U13 repeat: matches 1. 103 of consensus"
13807. 13838
repeat_region
/note="MER50 repeat: matches 573. 604 of consensus"
13850. 13900
repeat_region
/note="L1MB3 repeat: matches 6127. 6176 of consensus"
14441. 14924
repeat_region
/note="L2 repeat: matches 159. 652 of consensus"
15447. 15511
repeat_region
/note="L1M3 repeat: matches 5692. 5756 of consensus"
16041. 16314
repeat_region
/note="AluY repeat: matches 37. 310 of consensus"
16324. 16483
repeat_region
/note="MIR repeat: matches 53. 227 of consensus"
18337. 18491
repeat_region
/note="MIR repeat: matches 20. 188 of consensus"
20555. 20596
repeat_region

/note="21 copies 2 mer ct 78% conserved"
22105. 22178
repeat_region
/note="37 copies 2 mer at 70% conserved"
22111. 22190
repeat_region
/note="8 copies 10 mer atatgtatat 72% conserved"
22112. 22179
repeat_region
/note="17 copies 4 mer tata 72% conserved"
24126. 24193
repeat_region
/note="MIR repeat: matches 21. 88 of consensus"
24935. 25244
repeat_region
/note="MER58B repeat: matches 2. 330 of consensus"
25246. 25543
repeat_region
/note="AluY repeat: matches 1. 293 of consensus"
25837. 26126
repeat_region
/note="AluX repeat: matches 1. 310 of consensus"
26213. 26384
repeat_region
/note="FRAM repeat: matches 0. 173 of consensus"
27857. 27933
repeat_region
/note="MER3 repeat: matches 126. 187 of consensus"
27935. 28034
repeat_region
/note="HY3 repeat: matches 2. 101 of consensus"
28055. 28197
repeat_region
/note="MER3 repeat: matches 2. 144 of consensus"
29050. 29572
repeat_region
/note="MLT1F repeat: matches 1. 541 of consensus"
31088. 31172
repeat_region
/note="L1MC5 repeat: matches 7767. 7859 of consensus"
31271. 31579
repeat_region
/note="AluX repeat: matches 1. 312 of consensus"
32702. 33003
repeat_region
/note="L2 repeat: matches 1232. 1558 of consensus"
33033. 33281
repeat_region
/note="ORSL repeat: matches 203. 468 of consensus"
33324. 33538
repeat_region
/note="MIR repeat: matches 8. 219 of consensus"
33503. 34342
repeat_region
/note="L2 repeat: matches 1568. 2501 of consensus"
34360. 34515
repeat_region
/note="MLT1J repeat: matches 352. 516 of consensus"
37082. 37381
repeat_region
/note="AluSg repeat: matches 1. 299 of consensus"
37382. 37411
repeat_region
/note="10 copies 3 mer caa 93% conserved"
38652. 38928
repeat_region
/note="AluSg repeat: matches 1. 295 of consensus"
39931. 40014
repeat_region
/note="MER45C repeat: matches 821. 917 of consensus"
40029. 40159
repeat_region
/note="MER45C repeat: matches 433. 593 of consensus"
40163. 40705
repeat_region
/note="MER45C repeat: matches 1. 611 of consensus"
41895. 42201
repeat_region
/note="Aluub repeat: matches 1. 302 of consensus"
42805. 43107
repeat_region
/note="Aluuo repeat: matches 2. 308 of consensus"
43233. 43312
repeat_region
/note="8 copies 10 mer taataataa 71% conserved"
43241. 43300
repeat_region
/note="30 copies 2 mer ta 68% conserved"
43735. 44274
repeat_region
/note="L1MC1 repeat: matches 5776. 6332 of consensus"
44278. 44524
repeat_region
/note="AluSg repeat: matches 3. 313 of consensus"
45139. 45178
repeat_region
/note="20 copies 2 mer tt 77% conserved"
46887. 48493
repeat_region
/note="L2 repeat: matches 1011. 2710 of consensus"
49151. 49460
repeat_region
/note="AluSP repeat: matches 1. 310 of consensus"
50397. 50627
repeat_region
/note="Aluub repeat: matches 61. 294 of consensus"
51085. 51327
repeat_region
/note="MER8 repeat: matches 1. 219 of consensus"

```

repeat_region      51710..52001
                   /note="Alusx repeat: matches 9..306 of consensus"
repeat_region      52958..52979
                   /note="LIMB3 repeat: matches 5610..5630 of consensus"
repeat_region      52980..53283
                   /note="Alusx repeat: matches 1..304 of consensus"
repeat_region      53284..53622
                   /note="LIMB3 repeat: matches 5630..6183 of consensus"
repeat_region      53920..54162
                   /note="Alusx repeat: matches 1..243 of consensus"
repeat_region      55925..56208
                   /note="Alusx repeat: matches 1..294 of consensus"
repeat_region      56277..56643
                   /note="THE1B repeat: matches 1..364 of consensus"
repeat_region      57247..57409
                   /note="Sequence from overlapping clone AL136222. Assembly
confirmed by restriction digest data."
misc_feature        57422..57491
                   /note="35 copies 2 mer ca 75% conserved"
repeat_region      57423..57486
                   /note="16 copies 4 mer acac 76% conserved"
repeat_region      57431..57490
                   /note="6 copies 10 mer aacacacac 78% conserved"
repeat_region      58153..58459

Query Match:      100.0%; Score 14; DB 9; Length 123385;
Best Local Similarity 100.0%; Pred. No. 7,5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT	9
LOCUS	D90912
DEFINITION	D90912 128588 bp DNA BCT 04-JUL-2001 Synecocystis sp. PCC 6803 DNA, complete genome, section:14/27, 1719644-1848241.
ACCESSION	D90912 AB001339 BA000022
VERSION	D90912.1 GI:1653228
KEYWORDS	
SOURCE	.
ORGANISM	Synechocystis sp. PCC 6803 (strain:PCC6803) DNA. Synecocystis sp. PCC 6803 Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE	1 (bases 1 to 128598) Kaneke,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiura,M. and Tabata,S. Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome DNA Res. 2 (4), 153-166 (1995)
TITLE	96127529
JOURNAL	
MEDLINE	
REFERENCE	2 (bases 1 to 128598) Kaneke,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosewa,M., Sugiura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimo,S., Takeuchi,C., Wada,T., Metanabe,A., Yanada,M., Yasuda,M. and Tabata,S. Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions DNA Res. 3 (3), 109-136 (1996)
TITLE	97061201
JOURNAL	
MEDLINE	
REFERENCE	3 (bases 1 to 128598) Tabata,S. Direct Submission Submitted (26-JUN-1996) Satoshi Tabata, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 297-0812, Japan (E-mail:tabata@kazusa.or.jp, URL:http://www.kazusa.or.jp/cyano/ Tel.:81-438-52-3933(ext.2330), Fax:81-438-52-3934)
AUTHORS	Potential protein coding regions were assigned on the basis of
JOURNAL	
COMMENT	

similarity search of the ORFs and Genemark analysis.

FEATURES	source
gene	Location/Outils/ters 1. .128598 /organism="Synechocystis sp. PCC 6803" /strain="PCC6803" /db_xref="taxon:1148" /note="synonym:Synechocystis PCC6803"
CDS	complement(1. .933) /gene="s111263" complement(1. .933) /gene="s111263" /note="ORF_ID:s111263 hypothetical protein" /codon_start=1 /transl_table=1 /protein_id="BA18144.1" /db_xref="GI:1653229"
gene	/translation="MTARLARPYAVLSIGAAATMGKLGAVAYITGSVGLSDALEST VNLSAIVAFWALSTLATPADSEHPFGSKAEYSSGLEGAFTFVALGIGYSAYELET LSPALDIONALGATATAATPAINCTVAMITMRGKRLNTSIALRDSOHLMTDWTSTVG VYVAVALIFVTGWMEDLPILATLSGTFVMTITGHLLETTISLSDSLPRAODATITGS CFLEUDQGVFFHLQTRQGSQSFISFHVLPQHMVYQKHDLCEALETALERTITGS SRVTHLEPLEDPKSWQHPDEFPSPAPLNRDKPN"
CDS	complement(97. .1482) /gene="s111262" complement(97. .1482) /gene="s111262" /note="ORF_ID:s111262 hypothetical protein" /codon_start=1 /transl_table=1 /protein_id="BA18145.1" /db_xref="GI:1653230"
gene	/translation="MLRPLIANGKGFIRALENDGALAVYAPLEGEGYEGYORLRAN GRASISASARGLDVEAYIMQYHVRPAHLKKNITADGAVGPIYTFPAQIPAGYQENL PAQSGGLVMTLBSYILTSQTEIDLLSTFKRPAKLVLEMGDGRVERWOPLLDCLDA A"
CDS	complement(1707. .2363) /gene="tsf" complement(1707. .2363) /gene="tsf" /note="ORF_ID:s111261" /codon_start=1 /transl_table=1 /product="elongation factor ts" /protein_id="BA18146.1" /db_xref="GI:1653231"
gene	/translation="MAETTAQVLYHREKTAGAGMDCCKALKENEGDLSEIEMLRQK GIAADKKSQRTAGLVAHSHYIHGCRITGVAFYENCSTDPARDBRKDLVNDYAMQID AAGNVEVSVAADIPOENVAKEKEITEGRDDLQKPAPIKEKITYQGIIDKRLKLSLL DQPIKQDNLITELVKAIALAELENIQVRFIRENLGEGIEKAETNFAEEVAAAARG "
CDS	complement(2517. .3326) /gene="rps2" complement(2517. .3326) /gene="rps2" /note="ORF_ID:s111260" /codon_start=1 /transl_table=1 /product="30S ribosomal protein s2" /protein_id="BA18147.1" /db_xref="GI:1653232"
gene	/translation="MPVYSLADLLESCVHGCHQTRRNPNRDOYITTAANGVHITDLY QTALMEDAYEVSRSTTINGKRLFLVGTGRQACITISQEAQRCANHYNQHWLGGHILF NMEITRRVBDLKEALEALSSGSGIDRGRKEGSLRRELGLQKLYLGIGIKMRKLPV VVVVIDOREHNAIHECOKLGIPITAMDITCDPDVVDVPIIPANDAIRSLKILVGLLA DAIEGRHQPDVSDDYEEEPEDGIDGNLVEEAEEVEEAEEVAATPEA"
CDS	complement(3551. .3623) /product="tRNA-OTHER" 3809. .4927 /gene="s111336" 3809. .4927 /gene="s111336"
gene	/gene="s111336"
CDS	/gene="s111336"

```

/note="ORF_ID:s1r1336"
/codon_start=1
/transl_table=1
/product="H+/Ca2+ exchanger"
/protein_id="BA18148.1"
/db_xref="GI:1653233"
/translation="MSTKSKIFLVLLVFCPLSFRAHMLGCGETVVFILAGIAYPLAA
FMGTAETIEIAVIGIPNAGGLNATFGNATELILAYIALKEGLIEVKNATLIGSITGNL
LLVMGFVFLGLRKEQNFQPLAARLNASYTNLGVYAILPTALQYSTGVEETVLQ
NLISYAAVILIGVYLLSVFSMGTHAYLYDYGANMMPGLGDSPEPEEERK
NLMIMTGLVITGLVAVSESLIGLEVAETESIGLAFPGVIVLPTIGNAEHATA
VTVAMKDKMDLSKSYVMGSSSLQIAFPAVPVIVYGMALGQMDLNENFELVAVLVAV
LIVNISSDGTSNMLEGILLATYATVALAFAFFHPTLV"
comp_loment(4954. .5535)
/gene="s111258"
/complement(4954. .5535)
/note="ORF_ID:s111258"
/unknown_protein"
/codon_start=1
/transl_table=1
/protein_id="BA18149.1"
/db_xref="GI:1653234"
/translation="MIKNDITRAQAQCMTEPEPKLMROIIEGPAISGELSSFGD
IRLSAADFRIEIRHIGTVVDCKNPNPHLEATQLRHDSGYSFLLPHASVGLGVAVER
LAIPANVTVCIGKSTYARAGIILNLTPEGCMGCLTLERSSSSADCRITANEGLV
QLLFEGEDCDISYETRGRKYONQPSVTLPRV"
complement(5788. .6282)
/gene="s111257"
/complement(5788. .6282)
/complement(5798. .6282)
/gene="s111257"
/note="ORF_ID:s111257"
/codon_start=1
/transl_table=1
/product="transposase"
/protein_id="BA18150.1"
/db_xref="GI:1653235"
/translation="MLRELILYGSQAIVYIDESGFPAIOACIYAMSKKCKYVGDRO
GKRGVRENLVAGRRKGLDIPAWFTGSLNAEGFEGMLKYLPLPSIDISILMDNA
PIHRKTAIKELAKAGHEVLELPKYSPLNDLIEHDFALKRARMYADIDSLDEIIS
YCGV"
complement(6276. .6635)
/gene="s111256"
/complement(6276. .6635)
/gene="s111256"
/note="ORF_ID:s111256"
/codon_start=1
/transl_table=1
/product="transposase"
/protein_id="BA18151.1"
/db_xref="GI:1653236"
/translation="MAYSLDLRORVAVYIEAGKITEASKIYKIGKASIRYMLRVDL
SPTKVERHRRLDEWALKDYENPDARLIDRAKKFQVRSAVYVYALKKKKINKKKK
YVIEKGRNELSTIEC"
7034. .7656
/gene="s1r1338"
7054. .7656
/gene="s1r1338"
/note="ORF_ID:s1r1338"
/unknown_protein"
/codon_start=1
/transl_table=1
/protein_id="BA18152.1"
/db_xref="GI:1653237"
/translation="MSNORDYTLILIOKSGMSIIEPKSRMELVOESPLALARKD
OLDANGITYVTSKRRYDNVNASKEYQJFOENRPGGINTLAVLADALNLTORKK
SNAPTEGTLVITDGEPNDRSRVFEITIQASKCLDDELATISFMQIGMDPSATKFL
QALDDLMVEGAFDIVDTVTFDEMDLTLEVLNAIED"
7640. .8098
/gene="s1r1339"
7640. .8098
/gene="s1r1339"
/note="ORF_ID:s1r1339"

```

```

unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BA18153.1"
/db_xref="GI:1653238"
/translation="MLRIETRYTMMNSDSISLADLRKITYEPELTSSNPALIS
TSNEQHPHPTIDITLNEIKGCFKVNYSPTPKIOATFKOTSSEINFLKRLIDHROQ
ORQAIIIOKQOOWLEQDPLSGEIMWADLAKHVPSTLAAIALNLNNEST"
8127. .8663
/gene="s1r1340"
8127. .8663
/gene="s1r1340"
/note="ORF_ID:s1r1340"
/unknown_protein"
/codon_start=1
/transl_table=1
/protein_id="BA18154.1"

```

Query Match 100.0%; Score 14; DB 1; Length 128598;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cttgctgaacagg 14
Db 13060 CTTGCTGAACAGC 13073

RESULT 10
AL136222/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 150447)
Sims, S.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9211796.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: dj231G18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 143449 bases at least Q40
Consensus quality: 146211 bases at least Q30
Consensus quality: 147785 bases at least Q20
Insert size: 149147; sum-of-contigs
Insert size: 152905; 30.9% error; agarose-fp
Quality coverage: 3.73x in Q20 bases; sum-of-contigs Quality
coverage: 3.75x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

	*	1	10381:	contig of 10381 bp	in length.
	*	10382	10481:	gap of 100 bp	
	*	10482	22210:	contig of 11729 bp	in length
	*	22211	22310:	gap of 100 bp	
	*	22311	30167:	contig of 7857 bp	in length
	*	30168	30267:	gap of 100 bp	
	*	30268	32644:	contig of 2377 bp	in length
	*	32645	32744:	gap of 100 bp	
	*	32745	44240:	contig of 11496 bp	in length
	*	44241	44340:	gap of 100 bp	
	*	44341	47614:	contig of 12568 bp	in length
	*	47615	47714:	gap of 100 bp	
	*	47715	60282:	contig of 12568 bp	in length
	*	60283	60382:	gap of 100 bp	
	*	60383	103854:	contig of 45472 bp	in length
	*	103855	105954:	gap of 100 bp	
	*	105955	116556:	contig of 10602 bp	in length
	*	116557	116556:	gap of 100 bp	
	*	116557	124204:	contig of 7548 bp	in length
	*	124205	124304:	gap of 100 bp	
	*	124305	131202:	contig of 6888 bp	in length
	*	131203	131302:	gap of 100 bp	
	*	131303	133909:	contig of 2607 bp	in length
	*	133910	134009:	gap of 100 bp	
	*	134010	145780:	contig of 11771 bp	in length
	*	145781	145880:	gap of 100 bp	
	*	145881	150447:	contig of 4567 bp	in length.
FEATURES					
source					
			1..150447		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/chromosome="6"		
			/clone="RP1-231G18"		
			/clone_1fb="RPC1-1"		
			1..10381		
misc_feature			/note="assembly-fragment:00432		
			fragment_chain:1"		
misc_feature			10482..22210		
			/note="assembly-fragment:00888		
			fragment_chain:1"		
misc_feature			22311..30167		
			/note="assembly-fragment:00135		
			fragment_chain:1"		
misc_feature			30268..32644		
			/note="assembly-fragment:01177		
			fragment_chain:1"		
misc_feature			32745..44240		
			/note="assembly-fragment:01317		
			fragment_chain:1"		
misc_feature			44341..47614		
			/note="assembly-fragment:00959		
			fragment_chain:2"		
misc_feature			47715..60282		
			/note="assembly-fragment:00141		
			fragment_chain:2"		
misc_feature			60383..105854		
			/note="assembly-fragment:00975		
			fragment_chain:2"		
misc_feature			105955..116556		
			/note="assembly-fragment:00291		
			fragment_chain:2"		
misc_feature			116557..124204		
			/note="assembly-fragment:00297		
			fragment_chain:3"		
misc_feature			124305..131202		
			/note="assembly-fragment:00965		
			fragment_chain:3"		
misc_feature			131303..133909		
			/note="assembly-fragment:01066"		
misc_feature			134010..145780		
			/note="assembly-fragment:01208		
			fragment_chain:4"		
misc_feature			145881..150447		

```

/Note="assembly fragment:00924
fragment chain:4
clone_end:T7
vector_side:right"
BASE COUNT      41420 a 32488 c 33670 g 41560 t 1309 others
ORIGIN
Query Match      100.0%; Score 14; DB 2; Length 150447;
Best Local Similarity 100.0%; Prid. No. 7.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cttgcttgaacagg 14
|||||
Db 102675 CTTGCTTGACAGG 102662

RESULT 11
AC021608/c
LOCUS
DEFINITION
Homo sapiens clone RP11-233022, WORKING DRAFT SEQUENCE, 26
unordered pieces.
AC021608
AC021608.3 GI:7229829
HTGS: HTGS_PHASE1: HTGS_DRAFT.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 158469)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-233022
Unpublished
2 (bases 1 to 158469)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckler,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burrell,G., Castle,A.,
Choepele,T., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenster,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McNeheaters,R., Meldrim,J., Menues,L., Morlow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:5981873.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I5814
Center clone name: 233-O-22
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 140620 bases at least Q40
Consensus quality: 149963 bases at least Q30
Consensus quality: 153612 bases at least Q20

```

Insert size: 171000; agarose-fp
Insert size: 155969; sum-of-contigs
Quality coverage: 3.3 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1053: contig of 1053 bp in length
1054 1153: gap of 100 bp
1154 2231: contig of 1078 bp in length
2232 2331: gap of 100 bp
2332 3775: contig of 1444 bp in length
3776 3875: gap of 100 bp
3876 5598: contig of 1723 bp in length
5599 5698: gap of 100 bp
5699 7637: contig of 1939 bp in length
7638 7737: gap of 100 bp
7738 9294: contig of 1557 bp in length
9295 9394: gap of 100 bp
9395 11578: contig of 2184 bp in length
11579 11678: gap of 100 bp
11679 14096: contig of 2418 bp in length
14097 14196: gap of 100 bp
14197 17316: contig of 3120 bp in length
17317 17416: gap of 100 bp
17417 19700: contig of 2284 bp in length
19701 19800: gap of 100 bp
19801 22410: contig of 2610 bp in length
22411 22510: gap of 100 bp
22511 25738: contig of 3228 bp in length
25739 25838: gap of 100 bp
25839 29436: contig of 3598 bp in length
29437 29536: gap of 100 bp
29537 33121: contig of 3585 bp in length
33122 33221: gap of 100 bp
33222 38629: contig of 5408 bp in length
38630 38729: gap of 100 bp
38730 44943: contig of 6214 bp in length
44944 45043: gap of 100 bp
45044 51689: contig of 6646 bp in length
51690 51789: gap of 100 bp
51790 58196: contig of 6407 bp in length
58197 58296: gap of 100 bp
58297 64262: contig of 5966 bp in length
64263 64362: gap of 100 bp
64363 71550: contig of 7188 bp in length
71551 71650: gap of 100 bp
71651 82576: contig of 10926 bp in length
82577 82676: gap of 100 bp
82677 92884: contig of 10008 bp in length
92885 92784: gap of 100 bp
92785 104546: contig of 11762 bp in length
104547 104646: gap of 100 bp
104647 119207: contig of 14561 bp in length
119208 119307: gap of 100 bp
119308 134673: contig of 15366 bp in length
134674 134773: gap of 100 bp
134774 158469: contig of 23696 bp in length.

Location/Qualifiers

FEATURES

SOURCE

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-233022"
/clone_id="RPC1-11 Human Male BAC"
1. 1053
misc_feature
misc_feature
1154. 2231

/note="assembly_fragment"
2332. 3775
/note="assembly_fragment"
3876. 5598
/note="assembly_fragment"
5699. 7637
/note="assembly_fragment"
7738. 9294
/note="assembly_fragment"
9395. 11578
/note="assembly_fragment"
11679. 14096
/note="assembly_fragment"
14197. 17316
/note="assembly_fragment"
17417. 19700
/note="assembly_fragment"
clone_end:17
vector_side:left"
19801. 22410
/note="assembly_fragment"
22511. 25738
/note="assembly_fragment"
25839. 29436
/note="assembly_fragment"
29537. 33121
/note="assembly_fragment"
33222. 38629
/note="assembly_fragment"
38730. 44943
/note="assembly_fragment"
45044. 51689
/note="assembly_fragment"
51790. 58196
/note="assembly_fragment"
58297. 64262
/note="assembly_fragment"
64363. 71550
/note="assembly_fragment"
71651. 82576
/note="assembly_fragment"
82677. 92884
/note="assembly_fragment"
92785. 104546
/note="assembly_fragment"
104647. 119207
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
119308. 134673
/note="assembly_fragment"
134774. 158469
/note="assembly_fragment"
BASE COUNT 50039 a 32119 c 29755 g 44050 t 2506 others
ORIGIN

Query Match

Best local similarity 100.0%; Pred. No. 74e+02; Length 158469;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgctgaacagg 14
|||||

Db 17230 CTGCTGACACGG 17217

RESULT 12

AP002522 163526 bp DNA PIN 26-JAN-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0009603.
ACCESSION AP002522
VERSION AP002522.1 GI:8570076
KEYWORDS

SOURCE
ORGANISM
Oryza sativa (cultivar: Nipponbare) DNA, clone: P0009G03.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaristidae; Oryzaceae; Oryza.
REFERENCE
AUTHORS
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P0009G03
JOURNAL
Published only in Database (2000) In press
2 (bases 1 to 163526)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (14-JUN-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@db.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
tel: 81-298-38-7441, Fax: 81-298-38-7468)
Genes were predicted from the integrated results of the following:
GENSCAN 1.0, BLASTN 2.0, BLASTX 2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP 2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID. A gene with
identity or significant homology to a protein is classified based
on the protein name to indicate the homology level such as same
name, 'putative-' and '-like protein'. A gene without significant
homology to any protein but with EST homology (covering almost the
entire length) is classified as an 'unknown' protein. A gene
predicted with a gene prediction program is classified as a
'hypothetical' protein.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rgp.dna.affrc.go.jp/genomeseq.html.
location/Qualifiers
1..163526
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0009G03"
complement(join(4024..4697,4998..6318,6860..6954,
7585..7741))
/gene="P0009G03.1"
complement(join(4024..4697,4998..6318,6860..6954,
7585..7741))
/gene="P0009G03.1"
/note="contains ESTs AU031804(R1841), D24401(R1841)
unknown protein"
/codon_start=1
/protein_id="BAB03602.1"
/db_xref="GI:9663961"
/translation="MAQLRHODSAFYKELHGRKMSILQFGRRLRSTRMIDKK
QSGQSGDEKIVGSRRRGSTVPLKDETSVDDKHTETTKNKSRSKSSKSSGLG
SLILKLKYEGEKHKLMPVAKRLRLTSLIHYLENEVYLDDESANGSGSSIMPN
STDNIQATNATGDCNDNDSSTSLKRGDIHVKSQSHRSIMDGVLHKPYGHKSE
DYIKGLSRSASATGCDKDKISYTVAAKRPVNOEFRSRSRLTESLEKYSHLDSIS
NOSKRLTSSKSTRDHSIDAPVAMVGLQSIWTAENIVPEHALAPYBERITVGDV
DAAMDIEISHRKADSENIVLEIYIVDKKGVSTANICTPPLISEVIDISGHV
ATCDDQVHSTEDNMCTVHSKSEIDVPEEGTCCDDHNLSTEGAHHTKVSDEG
RISEHRTTSDKQIHSBVLKSRGTCVPSQEIARINLSCEQETDSFMSVLD
TPDDPASPVKYTLIDSSLKPRVLAHSDTDSADMDINTSMELSTAGLHKVNOE
SOSDELSLOADPKNEDGLIYKNIEMKSEFNEVLEFDMCSQNTIVLOEEDCOHYE
AAADPFEMESDOLLISDLTNEVLIDYKKYSVSKSLSRSSFDPRKPCVGNHAIKE
LMSKYSCHIDEPQSIETDITLSNDLSKNDHNVFQRDADHIGMLADFYDRLLTE
FTLQAKF"
join(8504..8511,9977..9160)
/gene="P0009G03.2"
join(8504..8511,9977..9160)
CDS

/gene="P0009G03.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB03603.1"
/db_xref="GI:9663962"
/translation="MDRLNARRCHOEPTKPAKKPPFISANSPSISDDACSRNDEKK
GGVRFEGVLVLPDGGSW"
complement(join(12593..12661,12739..13502,15691..15730,
15945..16013))
/gene="P0009G03.3"
complement(join(12593..12661,12739..13502,15691..15730,
15945..16013))
/gene="P0009G03.3"
complement(join(12593..12661,12739..13502,15691..15730,
15945..16013))
/note="contains ESTs AU031642(R0106), D23767(R0106)"
/codon_start=1
/product="putative glycoprotein"
/protein_id="BAB03604.1"
/db_xref="GI:9663963"
/translation="MKVSIVSRSGREVGVGLKDSKAVADLODAIYAKTKYYPAR
ORLTPIOPGKSGKPVLSAKASLSYCEKSGSLIVVRCFKIDGPQVFSYTLFFELIG
PLIITPMFYILPYKYKFCYEGERNVHPQVITYMYWCPIYFKRIETPFVHPSHATS
PVSNNFNCVATWIGATYATYCNHPLTPVSELQKIGFGILCOINFFCHILLR
NLRSPGNGCYOIPRGFLNIVTCANYTTEIYQMDGFNATQTVAGYIFLVVAASIMT
NMALKHRLKASTLFDGEGRPKPYRRWVILPEFL"
join(17137..17202,17954..18253,19340..19415,19517..19575,
19685..19825,20367..20422,20509..20585,20684..20729,
22725..22851,22943..23072)
/gene="P0009G03.4"
join(17137..17202,17954..18253,19340..19415,19517..19575,
19685..19825,20367..20422,20509..20585,20684..20729,
22725..22851,22943..23072)
/gene="P0009G03.4"
/codon_start=1
/product="putative geranylgeranyl-diphosphate
geranylgeranyltransferase I beta chain"
/protein_id="BAB03605.1"
/db_xref="GI:9663964"
/translation="MADPSRSYRIKPNSEMYTSRKHQRPPEPPPCRALILQIR
TSGVILSGRDGSPASERGPAMAEPEDEPARRHVLEFLEMASLDPADYASQEVN
HLTIAYFAVAGLSLRELDSPVKKDIAMKILSFQVHPRTDNLNGQFPGFCGRTO
PEPTNMKDPCHNGSHASTYSALATIKIYGDLANIDNVLSKRNLODPGSEMPPT
HIGAEITDRFYCAATICSMLKDMTKGEKAKOYILSCSYDGGFGLVPGSESHGLA
FLGSSSLLAAECASMTGVGSAVRILYSVNMVAFGMAQVLYVATTTLESQHEA
VYAAVERPQAPQAPALFVWFC"
join(23889..23928,24196..24284,24674..24727,24878..24948,
25052..25133,25238..25334,25455..25525)
/gene="P0009G03.5"
join(23889..23928,24196..24284,24674..24727,24878..24948,
25052..25133,25238..25334,25455..25525)
/note="contains ESTs AU089745(S14080), AU097665(C30644)
unknown protein"
/codon_start=1
/protein_id="BAB03606.1"
/db_xref="GI:9663965"
/translation="MALIYAOLIHGLVGRSPVSAATLPQIGSRILFMWGLMSPEPT
HSHILVTSIYVSWSTETETIRTSFRCMKETFGCAPSMLRLKRTSMWLKPTGISSEVG
LIYILPYMKATEKCYCLRMPKNWNSPDSVASIISLAVYVGSHPMTYMAORRKA
LAKAKAA"
join(27342..27412,27505..27616,27769..27787,28055..28143,
28491..28544,28754..28863,28990..29071,29170..29255,
29439..29510,29748..29893,30841..30902)
/gene="P0009G03.6"
join(27342..27412,27505..27616,27769..27787,28055..28143,
28491..28544,28754..28863,28990..29071,29170..29255,
29439..29510,29748..29893,30841..30902)
/gene="P0009G03.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB03607.1"
/db_xref="GI:9663966"
/translation="MAADGSVLRRLYFSYNNVVFIGNAOVSLSMILALDDGHEAVY
AATERRHLFAQPAAIMETLHSTILGLVRSPVSTLPQITGRILFTWGLMSPEPTSHI

```

gene
    CDS
        34694..34747,34987..35057,35181..35262,35355..35451,
        35876..35944)
        /gene="P0009603.7"
        /join(32812..32882,33147..33264,33378..33396,34095..34183,
        34694..34747,34987..35057,35181..35262,35355..35451,
        35876..35944)
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="BAB03608.1"
        /db_xref="GI:9663967"
        /translation="MADGSMYRLYLSTYVAVFAGMVAOVLCYVTLALDQKHEAVY
        AAIERPLTQTAALTEVLIHSTYGLVSRPSSLPITGSLFTTKGLNSPEPTOS
        YIFVSLICWITIEVTRSTFGKESGFTPSLWMLRSTFICFVGVVSEITCLA
        YTVLPKASEKYLCLMPKMFNFYVAVFMAFYVPCMKHPLNEVDLEAFTNC
        SSALGTMPQSVLRFKDL"
        complement(join(37944..38212,38801..39038,39214..39345))
        /gene="P0009603.8"
        complement(join(37944..38212,38801..39038,39214..39345))
        /note="probably inactive due to frameshifts and stop codon
        in CDS and no initiation/termination codons
        pseudogene, hypothetical protein"
        /codon_start=1
        /pseudo
        complement(join(40186..41271,41379..41555,41734..42031,
        42201..42399,42444..42507))
        /gene="P0009603.9"
        complement(join(40186..41271,41379..41555,41734..42031,
        42201..42399,42444..42507))
        /gene="P0009603.9"
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="BAB03609.1"
        /db_xref="GI:9663968"
        /translation="MAVKDRFTGNPGRLLVLTSGTSGEGRDQVSAEYGCPCAG
        FTAOEGTAHLLRGCGEILAMCGVGEVGIYTIISPAVSCSGMLGILQORVAL
        DSDMLILINCFATFAPYCEMPALCKRTSFLMLYIPATLIFRDLSTVGSITOC
        SLFPRIVELVLSRLREVGVAVALKNSGRISGCDTALDRPRLALAGGGS
        LSGKSSRPHPDQDQOQSVIRPSRETIHOVQASLCPNLSNLLVLVMAVSDP
        FLGGALSTDIKLGQNYQEWELASRLISGASHLTDNPEKTLDAIKAMKT
        ADDRVMGTFMSVEVPIRMSFRDSTAKEMDYLKORYTOSGALRSLDNLHLQO
        ODOSIEEFYNAFTLSQLEAOTPGASGACQCKAREKHODENLVYQPVMSLSQFES
        IRVQLGRPTMAELADLIAETRLCSLDTTPPYVTNNMAAPORVAPMGGIP
    
```

Query Match 100.0%; Score 14; DB 8; Length 163526;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgctgacagg 14
 |||||

Db 6826 cttgctgacagg 6839

RESULT 13
 AC007876
 LOCUS AC007876 164879 bp DNA PRI 14-OCT-2000
 DEFINITION Homo sapiens BAC clone RP11-65G9 from Y, complete sequence.
 AC007876
 ACCESSION
 VERSION
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 164879)
 AUTHORS
 TITLE
 Toward a complete human genome sequence

JOURNAL
 MEDLINE
 99063792
 2 (bases 1 to 164879)
 Hou, S., Kalicki, J., and Stumpf, J.
 The sequence of Homo sapiens BAC clone RP11-65G9
 Unpublished
 3 (bases 1 to 164879)
 Waterston, R.H.
 Direct Submission
 Submitted (19-JUN-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 164879)
 Waterston, R.H.
 Direct Submission
 Submitted (19-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 164879)
 Waterston, R.H.
 Direct Submission
 Submitted (21-DEC-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 6 (bases 1 to 164879)
 Waterston, R.H.
 Direct Submission
 Submitted (03-MAY-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Dec 20, 1999 this sequence version replaced gi:5103904.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0065G09

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a
 collaboration between the Human Chromosome Y Mapping Project
 (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
 and David C. Page at the Whitehead Institute for Biomedical
 Research, Cambridge MA) and the Washington University Genome
 Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPI1 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
 Tareno, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute

(http://bacpac.med.buffalo.edu)
 VECTOR: PBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-33914; the clone sequenced
 to the right is RP11-470K20, 200 bp overlap. Actual start of this
 clone is at base position 1 of RP11-65G9; actual end is at base
 position 164685 of RP11-65G9.
 Location/Qualifiers

FEATURES
 SOURCE
 1. 164879
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="Y"
 /map="Y"
 /clone_lib="RP11-65G9"
 /clone_lib="RP11-65G9"
 1. 826
 /rpt_family="Retroviral"
 repeat_region
 1036. 1225
 /rpt_family="BSR/Beta"
 repeat_region
 1314. 1722
 /rpt_family="BSR/Beta"
 repeat_region
 3013. 3509
 /rpt_family="Retroviral"
 repeat_region
 3510. 4722
 /rpt_family="Retroviral"
 repeat_region
 5126. 5471
 /rpt_family="Retroviral"
 repeat_region
 5705. 6261
 /rpt_family="Retroviral"
 repeat_region
 9786. 9899
 /rpt_family="Retroviral"
 repeat_region
 9973. 10041
 /rpt_family="BSR/Beta"
 repeat_region
 10078. 10817
 /rpt_family="BSR/Beta"
 repeat_region
 10965. 11127
 /rpt_family="MER4-group"
 repeat_region
 11143. 11446
 /rpt_family="BSR/Beta"
 repeat_region
 11500. 11636
 /rpt_family="BSR/Beta"
 repeat_region
 12754. 12945
 /rpt_family="BSR/Beta"
 repeat_region
 13488. 13584
 /rpt_family="BSR/Beta"
 repeat_region
 13834. 13921
 /rpt_family="BSR/Beta"
 repeat_region
 14328. 14675
 /rpt_family="BSR/Beta"
 repeat_region
 15270. 15578
 /rpt_family="BSR/Beta"
 repeat_region
 16156. 16360
 /rpt_family="BSR/Beta"
 repeat_region
 16388. 16411
 /rpt_family="BSR/Beta"
 repeat_region
 17770. 18058
 /rpt_family="(TG)n"
 repeat_region
 18116. 18288
 /rpt_family="BSR/Beta"
 repeat_region
 18361. 18550
 /rpt_family="BSR/Beta"
 repeat_region
 18943. 19145
 /rpt_family="BSR/Beta"
 repeat_region
 19304. 19447
 /rpt_family="BSR/Beta"
 repeat_region
 19548. 19578
 /rpt_family="BSR/Beta"
 repeat_region
 19625. 225639
 /rpt_family="(TTTC)n"
 repeat_region
 25651. 25763
 /rpt_family="L1"
 repeat_region
 25771. 25800
 /rpt_family="MER4-group"
 repeat_region
 /rpt_family="(TTTA)n"

repeat_region 25801. 26066
 /rpt_family="Alu"
 repeat_region 2667. 27093
 /rpt_family="MER4-group"
 repeat_region 27088. 27139
 /rpt_family="MER4-group"
 repeat_region 27140. 27156
 /rpt_family="Retroviral?"
 repeat_region 27158. 27244
 /rpt_family="MER4-group"
 repeat_region 27274. 27524
 /rpt_family="Retroviral"
 repeat_region 27654. 28218
 /rpt_family="Retroviral"
 repeat_region 28347. 28674
 /rpt_family="(GGAA)n"
 repeat_region 28712. 28751
 /rpt_family="MER4-group"
 repeat_region 28803. 29052
 /rpt_family="Retroviral?"
 repeat_region 28909. 29081
 /rpt_family="Retroviral"
 repeat_region 29125. 30702
 /rpt_family="MER4-group"
 repeat_region 30901. 31123
 /rpt_family="BSR/Beta"
 repeat_region 31306. 31467
 /rpt_family="BSR/Beta"
 repeat_region 31520. 31928
 /rpt_family="BSR/Beta"
 repeat_region 32501. 32807
 /rpt_family="BSR/Beta"
 repeat_region 33142. 33482
 /rpt_family="Alu"
 repeat_region 33685. 33746
 /rpt_family="Retroviral"
 repeat_region 33750. 34079
 /rpt_family="Retroviral"
 repeat_region 34069. 34217
 /rpt_family="Retroviral"
 repeat_region 34217. 34789
 /rpt_family="Retroviral"

Query Match 100.0%; Score 14; DB 9; Length 164879;
 Best Local Similarity 100.0%; Pred. NO. 7.4e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgctgaacag 14
 |||1111111111
 Db 160726 CTTGCTGACACG 160739

RESULT 14
 AC080161/C
 LOCUS Homo sapiens chromosome 11, clone RP11-210C6, WORKING DRAFT
 DEFINITION
 SEQUENCE, 19 unordered pieces.
 ACCESSION AC080161
 VERSION AC080161.3 GI:11136856
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 166955)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 166955)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (27-SEP-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,

COMMENT

MO 63108, USA
On Nov 10, 2000 this sequence version replaced gi:11094872.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0210C06
Summary Statistics

Sequencing vector: M13: 1008
Sequencing vector: plasmid: 08
Chemistry: Dye-terminator Big Dye: 08 of reads
Chemistry: Dye-terminator Big Dye: 08 of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 156634 bases at least Q40
Consensus quality: 159676 bases at least Q30
Consensus quality: 161074 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 165155; sum-of-ctrls
Quality coverage: 4.34 in Q20 bases; sum-of-ctrls
Quality coverage: 4.62 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
4216: contig of 4216 bp in length
4217
4317
8023: contig of 3707 bp in length
8024
8124
12369: contig of 4246 bp in length
12370
12469: gap of unknown length
12470
18357: contig of 588 bp in length
18358
18457: gap of unknown length
18458
24671: contig of 6214 bp in length
24672
24771: gap of unknown length
24772
31294: contig of 6523 bp in length
31295
31394: gap of unknown length
31395
39432: contig of 8038 bp in length
39433
39532: gap of unknown length
39533
49624: contig of 10092 bp in length
49625
49724: gap of unknown length
49725
62485: contig of 12761 bp in length
62486
62585: gap of unknown length
62586
78199: contig of 15614 bp in length
78200
78299: gap of unknown length
78300
94935: contig of 16636 bp in length
94936
95035: gap of unknown length
95036
113050: contig of 18015 bp in length
113051
113150: gap of unknown length
113151
150215: contig of 37065 bp in length
150216
150315: gap of unknown length
150316
151874: contig of 1559 bp in length
151875
151974: gap of unknown length
151975
154742: contig of 2768 bp in length
154743
154842: gap of unknown length
154843
156956: contig of 2114 bp in length
156957
157056: gap of unknown length
157057
160139: contig of 3083 bp in length
160140
160239: gap of unknown length
160240
163106: contig of 2867 bp in length
163107
163206: gap of unknown length
163207
166955: contig of 3749 bp in length.

FEATURES

SOURCE

1.166955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"

misc_feature
1.4216
/clone="RP11-210C6"
misc_feature
4317.8023
/note="assembly_name:Contig10"
misc_feature
8124.12369
/note="assembly_name:Contig11"
misc_feature
12470.18357
/note="assembly_name:Contig12"
misc_feature
18458.24671
/note="assembly_name:Contig13"
misc_feature
24772.31294
/note="assembly_name:Contig14"
misc_feature
31395.39432
/note="assembly_name:Contig15"
misc_feature
39533.49624
/note="assembly_name:Contig16"
misc_feature
49725.62485
/note="assembly_name:Contig17"
misc_feature
62586.78199
/note="assembly_name:Contig18"
misc_feature
78300.94935
/note="assembly_name:Contig19"
misc_feature
95036.113050
/note="assembly_name:Contig20"
misc_feature
113151.150215
/note="assembly_name:Contig21"
misc_feature
150316.151874
/note="assembly_name:Contig22"
misc_feature
151975.154742
/note="assembly_name:Contig4"
misc_feature
154843.156956
/note="assembly_name:Contig5"
misc_feature
157057.160139
/note="assembly_name:Contig6"
misc_feature
160240.163106
/note="assembly_name:Contig7"
misc_feature
163207.166955
/note="assembly_name:Contig8"
misc_feature
163207.166955
/note="assembly_name:Contig9"
BASE COUNT 48590 a 31540 c 31064 g 53953 t 1808 others
ORIGIN
Query Match 100.0% Score 14: DB 2: Length 166955;
Best Local Similarity 100.0% Pred. NO. 7.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ctctctgaacagg 14
Db 76378 CTGCTGACAGG 76365
RESULT 15
AP002540
LOCUS AP002540 167029 bp DNA PLN 28-APR-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0434B04.
ACCESSION AP002540
VERSION AP002540.2 GI:13872872
KEYWORDS
SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: P0434B04.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (sites)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0434B04
JOURNAL Published Only in Database (2000) In press

26700..26668,26944..26985)
/gene="P0434B04.7"
complement(join(24107..24112,25628..25830,25915..26139,
26700..26668,26944..26985))
/gene="P0434B04.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB43986.1"
/db_xref="GI:13872879"
/translation="MAGREDAAAAAAGEVEEESDEEVYVEEEEGEGEGGAGE
TEGSGSGEKERKSSPRSCGSSPSPGEEVSEDEVQVGVLEKRRKATVPKQIIS
KIEIPHKQISEVOTPHKQISEVETPHQQISEVEIKSEVQALVLAQDSQRIQEN
ENLRQLALKTRELEHEENOKLRQLLELKNQIDIESLKKQNDLKAIEIYYKTS"
complement(join(28686..28919,29036..29230,29355..29633,
29736..29825,30033..30214,31141..32245))
/gene="P0434B04.8"
complement(join(28686..28919,29036..29230,29355..29633,
29736..29825,30033..30214,31141..32245))
/gene="P0434B04.8"
/note="contains ESTs
C27232(C51376),AU030178(E50721),AU030179(E50721)
unknown protein"
/codon_start=1
/protein_id="BAB43987.1"
/db_xref="GI:13872880"
/translation="MDLYMDYIDPYEEAEAAEAAGLTSAAAEDSDGDESD
SEASDYEEEREGYGLRSGNHRNPDPGFPFCPCGKKKODYKIKDLLOHADGIGASS
KRRHGREARAFARFVPTDPSFADLIAINGIATTAIDISNSAAGAKERADANG
DTDSASVAAVAAGNRREERFAMPNPGVYLAAGGFNAEDFADRYAAGGLDVPFLVY
DDAEGADSFALVRFANGWGGFDALTLENHFNKRLCKNENDARGSGADAVKGDGDT
DVKYGWVARGEDNAGNVGRYLARKHTNLTTIDEVSKSESEKSGKVAIILATQIEAK
NRFLODLETKKATELSIARLEEDNRKLDHAYNEEMRNLHRRARDYALRYQDNENLK
LELDSRRRELNSRAKOLEKITAENASDRKLDQOKOKAKDQNSELELASTEORADE
VMKIVEDOKREKEDVLAAMLQLEKLEHKQOLEEVPRLNGTLQVMKHLGGDDGDTH
DKMKLSERLEHEKREKLELSGELVKKERESNDELQERKELINGLEDMLGRTAIGI
KRKGELEKRPQNAQCKRRYGNDDYETKAELVSSMQEIKPMSWHPYKLTITEVDDDD
TKRLHLIEYGGDVCNAVKTALSEVNEYNASGRYVPEPLMNERKGRKATMKELVLYIF
GQIETTSKRRRP"
complement(join(34107..34436,34551..34658,34762..34845,
gene

Query Match 100.0%; Score 14; DB 8; Length 167029;
Best Local Similarity 100.0%; Pred. No. 7,4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcttgacacag 14
|||||
Db 150924 CTGCTTGACACAG 150937

Search completed: December 8, 2001, 10:51:23
Job time: 4948 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:21:33 ; Search time 2889.21 Seconds

(without alignments)
52.070 Million cell updates/sec

Title: US-09-508-147-10

Perfect score: 14

Sequence: 1 ctgtcgtgacagc 14

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	14	100.0	231 10	AM265067 xg60904.x
2	14	100.0	306 10	BR456636 UT-M-B21-
3	14	100.0	330 10	BB312230 BB312230
4	14	100.0	428 10	AA467271 v106d05.r
5	14	100.0	435 13	A2823498 2M0097E15
6	14	100.0	438 10	A1154122 ud6f02.r
7	14	100.0	462 11	BG510135 sac63806.
8	14	100.0	479 10	AM133355 se17h03.y
9	14	100.0	479 13	A2072421 RPT-23-4
10	14	100.0	485 10	AA670946 vp91h01.r
11	14	100.0	486 13	AO873059 v58E6 mtN
12	14	100.0	501 10	AM032968 EST276527

13	14	100.0	501 11	BF723243	BF723243 mab29f05.
14	14	100.0	511 13	A2036425	A2036425 RPT-23-3
15	14	100.0	513 13	A2249093	A2249093 RPT-23-3
16	14	100.0	522 13	A2477416	A2477416 1M0296G21
17	14	100.0	527 10	BE689712	BE689712 uw57905.y
18	14	100.0	541 13	A2166156	A2166156 SP_0083.B
19	14	100.0	604 13	AO503365	AO503365 v57G11.mt
20	14	100.0	635 10	BE293976	BE293976 601173154
21	14	100.0	641 11	BF327660	BF327660 RC2-BN012
22	14	100.0	734 11	BF214304	BF214304 601848529
23	14	100.0	735 13	AO162543	AO162543 m9xb0013K
24	14	100.0	767 11	BF638179	BF638179 NF025F04P
25	14	100.0	769 11	BI221887	BI221887 602938415
26	14	100.0	769 11	BE822897	BE822897 GM700019A
27	14	100.0	826 13	CNS02X00	AL217767 Tetradon
28	14	100.0	835 11	BG399678	BG399678 602441543
29	14	100.0	879 11	BG469900	BG469900 602533360
30	14	100.0	930 11	BG747368	BG747368 602704843
31	14	100.0	933 13	CNS05FPQ	AL335303 Tetradon
32	14	100.0	935 13	CNS02UPB	AL214397 Tetradon
33	14	100.0	940 13	CNS024FJ	AL180712 Tetradon
34	14	100.0	975 13	CNS04KHO	AL294837 Tetradon
35	14	100.0	997 13	CNS02WBP	AL216862 Tetradon
36	14	100.0	1035 13	CNS03J4V	AL246424 Tetradon
37	14	100.0	1064 13	CNS04FOR	AL287748 Tetradon
38	14	100.0	1084 13	CNS045NM	AL275611 Tetradon
39	13.6	97.1	914 13	CNS02YTD	AL220090 Tetradon
40	13	92.9	50 10	AU106058	AU106058 AU106058
41	13	92.9	98 10	AA138090	AA138090 m97e07.r
42	13	92.9	109 10	AA310031	AA310031 EST180877
43	13	92.9	140 10	AW751523	AW751523 MR1-CR005
44	13	92.9	163 13	A2759068	A2759068 1M0551E20
45	13	92.9	164 13	AO336380	AO336380 RPT-11-3

ALIGNMENTS

RESULT 1
LOCUS AM265067
DEFINITION xg60904.x1 NCI-CGAP Co22 Homo sapiens cDNA clone IMAGE:2755062 3'
similar to SW:Y0JM.BACSU P54550 PROBABLE NMDH-DEPENDENT FLAVIN
OXIDOREDUCTASE Y0JM ; mRNA sequence.

ACCESSION AM265067
VERSION AM265067.1 GI:6641883
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 231)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

unknown library type
Possible reversed clone: polyT not found

Seq primer: -40UP from Glibco
High quality sequence stop: 227.

Location/Qualifiers

FEATURES

source
1..231
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2755062"
/clone_id="NCI CGAP Co22"
/tissue_type="colonic adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pAMP10; cDNA made by oligo-dT
priming. Non-directionally cloned into the upc sites of
pAMP10. Size-selected on agarose gel, average insert

size 500 bp. Primary library: non-amplified. cDNA
Library Preparation: David B. Krizman, Ph.D (NCI).
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 40 a 72 c 67 g 52 t

ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cttgctgaacagg 14
|||||
Db 117 CTTGCTGAACAGG 130

RESULT 2

BF456636

LOCUS 306 bp mRNA EST 01-DEC-2000
DEFINITION UI-M-B21-bkg-f-12-0-UI.s1 NIH_BMAP_MHI2.S1 Mus musculus cDNA clone

ACCESSION BF456636

VERSION BF456636.1 GI:11522805

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 306)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890

Email: mest@mail.nih.gov

Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements

The following repetitive elements were found in this cDNA sequence:
14-53, >AT-richlow_complexity

Seq primer: M13 Forward

POLYA-No.

FEATURES Location/Qualifiers

SOURCE

1..306
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-B21-bkg-f-12-0-UI"
/clone_1db="NIH_BMAP_MHI2.S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_MHI2.S1 library is a subtracted library derived
from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT 70 a 83 g 76 t

ORIGIN

Query Match 100.0%; Score 14; DB 11; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cttgctgaacagg 14
|||||
Db 43 CTTGCTGAACAGG 56

RESULT 3

BB312230

LOCUS 330 bp mRNA EST 10-JUL-2000
DEFINITION BB312230 RIKEN full-length enriched, adult male corpora

ACCESSION BB312230

VERSION BB312230.1 GI:9012935

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 330)

AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

,P., Endo,T., Fukuda,S., Fukunishi,Y., Hata,A., Hayatsu,N.,
Hirozane,T., Horii,F., Ishii,Y., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamataka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

,M., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Konno,H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoke,S., Sasaki

,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermotabilization and thermocatalytic of thermolabile enzymes by

tRNAse and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itch,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki

,Y. and Hayashizaki,Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES Location/Qualifiers

SOURCE

1..330
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B230333B10"
/clone_1db="RIKEN full-length enriched, adult male corpora
quadrigenina"
/sex="male"
/tissue_type="corpora quadrigenina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was

with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 120 a 86 c 72 g 157 t
ORIGIN

Query Match 100.0%; Score 14; DB 13; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cttgcttgacagc 14
|||||
Db 270 CTTGCTTGACAGC 257

RESULT 6
A1154122 438 bp mRNA EST 30-SEP-1998
LOCUS ud76f02.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1476795 5', mRNA sequence.
ACCESSION A1154122 GI:3682591
VERSION A1154122.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 438)
Marré, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucab, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HM Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:925151
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 438.

FEATURES

source
1..438
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1476795"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 101 a 108 c 117 g 112 t
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 cttgcttgacagc 14
|||||
Db 309 CTTGCTTGACAGC 322

RESULT 7
BG510135 462 bp mRNA EST 28-MAR-2001
LOCUS sac63a06.y1 Gm-cl072 glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-cl072-11 5' similar to TR:095TF2 095TF2 HYPOTHETICAL 54.4 KO
PROTEIN. ; mRNA sequence.
ACCESSION BG510135
VERSION BG510135.1 GI:13480792
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 462)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna,
A., Bolla, B., Marré, M., Hillier, L., Kucab, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
R., Ritter, F., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
vector to vector length read
vector to vector length is 475
High quality sequence stop: 421.

FEATURES

source
1..462
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl072-11"
/clone_lib="Gm-cl072"
/tissue_type="seedlings induced for symptoms of SDS
(Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 2-3 week old seedlings that were induced for symptoms
of SDS (Sudden Death Syndrome) disease by the
translocation of culture filtrate of *Fusarium solani* f.
sp. glycines (Plant Cell Report 18:375-380). Cultivar PI
567374 is partially resistant to the disease SDS. Plant
tissue (expanded leaves, folded leaves, and new shoots)
were collected at 1, 6, 24, and 48 hrs. after inoculation
and their mRNA pooled equally for cDNA construction. The
library was prepared using the Stratagene pBluescript II
SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA insert is protected
from XhoI digestion via methylation during first strand


```

source
1..479
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-437G7"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      135 a      96 c      123 g      125 t

```

ORIGIN

Query Match 100.0%: Score 14: DB 13: Length 479;
 Best Local Similarity 100.0%: Pred. No. 1.2e+03;
 Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 ctgcttgacag 14
 |||
 DB 392 CTTGCTTGACAG 405

RESULT 10

AA670946 485 bp mRNA EST 25-NOV-1997
 LOCUS vp91h01.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
 DEFINITION clone IMAGE:1092145 5', mRNA sequence.

ACCESSION AA670946
 VERSION AA670946.1 GI:2643025

SOURCE

EST.
 house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 485)
 Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:598377

Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 440.

FEATURES

SOURCE

Location/Qualifiers

1..485
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="1092145"
 /clone_lib="Stratagene mouse diaphragm (#937303)"
 /tissue_type="diaphragm"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1:
 EcoRI, Site 2: XhoI; Cloned unidirectionally from mRNA
 prepared from diaphragm muscle. Primer: Oligo dT. Average
 insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
 sequence: 5' GAATTCGACAG 3' ~3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3'."
 BASE COUNT 104 a 113 c 138 g 130 t
 ORIGIN

Query Match 100.0%: Score 14: DB 10: Length 485;
 Best Local Similarity 100.0%: Pred. NO. 1.2e+03;
 Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

OY 1 ctgcttgacag 14
 |||
 DB 241 CTTGCTTGACAG 254

RESULT 11

A0873059

LOCUS A0873059 486 bp DNA GSS 08-NOV-1999
 DEFINITION V58E6 mtN-3xHA/LacZ Insertion Library, strain AB972 Saccharomyces
 cerevisiae genomic 5', DNA sequence.

ACCESSION A0873059
 VERSION A0873059.1 GI:6285303

KEYWORDS

GSS.

SOURCE

baker's yeast.
 ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE

1 (bases 1 to 486)

Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 deSilva, S. A., Cheung, K. H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Heger, K., Miller, P., Roeder, G. S. and Snyder, M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption

TITLE

Unpublished (1999)

JOURNAL

Contact: Kumar A

COMMENT

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161

Email: anuj.kumar@yale.edu
 te of mtN-3xHA/LacZ insertion.
 Seq primer: GGCCCTCTCTTGACAGTAC
 Class: transposon-tagged.

FEATURES

SOURCE

Location/Qualifiers

1..486
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtN-3xHA/LacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS56-Sal; A yeast genomic DNA library was
 prepared in pHS56-Sal; genomic DNA was size-fractionated
 (DNA of roughly 2-3 kb in length) prior to cloning. This
 library was subsequently mutagenized with a mtN-3xHA/LacZ
 multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 181 a 90 c 91 g 124 t
 ORIGIN

Query Match 100.0%: Score 14: DB 13: Length 486;
 Best Local Similarity 100.0%: Pred. NO. 1.2e+03;
 Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

OY 1 ctgcttgacag 14
 |||
 DB 172 CTTGCTTGACAG 185

RESULT 12

AM032968 501 bp mRNA EST 18-MAY-2001
 LOCUS EST276527 tomato callus, TAMU Lycopersicon esculentum cDNA clone
 DEFINITION CLC17N7, mRNA sequence.

ACCESSION AM032968
 VERSION AM032968.1 GI:5891724

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon

TITLE

1 (bases 1 to 501)

AUTHORS

Alcala, J., Vredalov, J., White, R., Matern, A. L., Vision, T., Holt, I. E.,
 Liang, F., Upton, D., Craven, M. B., Bowman, C. L., Ahn, S., Romling,
 C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.

TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source location/Qualifiers
1. 501
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC17N7"
/clone_lib="tomato callus, TAMU"
/tissue_type="Callus"
/dev_stage="25-40 days old"
/lab_host="XL1-Blue MRF"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI; supplier: GibcoBRL laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus Est Library"

BASE COUNT 167 a 102 c 79 g 153 t
ORIGIN

Query Match 100.0%; Score 14; DB 10; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
Db 161 CTTGCTTGACAGG 148

RESULT 13
BF723243 501 bp mRNA EST 03-JAN-2001
LOCUS mab29f05.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
DEFINITION IMAGE:3971841 5', mRNA sequence.
ACCESSION BF723243
VERSION BF723243.1 GI:12024245
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 501)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Other ESTs: mab29f05.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbsr@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at:
<http://iml1.gov/image/html/lresources.shtml>
MGI:1471873
Seq primer: -40RP from Gibco
High quality sequence stop: 472.

FEATURES
source location/Qualifiers
1. 501
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3971841"

/clone_lib="Soares_NMEBA_branchial_arch"
/tissue_type="branchial arches"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTACCAATCGAGAGCGAGCGCGACATTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 113 a 117 c 136 g 135 t
ORIGIN

Query Match 100.0%; Score 14; DB 11; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
Db 13 CTTGCTTGACAGG 26

RESULT 14
A2036425 511 bp DNA GSS 01-MAR-2000
LOCUS RPCI-23-351D22.TJ RPCI-23 Mus musculus genomic clone RPCI-23-351D22
DEFINITION A2036425
ACCESSION A2036425
VERSION A2036425.1 GI:7123871
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 511)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akintet, B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kiol,M., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderinframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/db/bac_ends/mouse/bac_end_intro.html
Plate: 351 row: D column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES
source location/Qualifiers
1. 511
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-351D22"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies).
 BASE COUNT 139 a 101 c 132 g 139 t
 ORIGIN

Query Match 100.0%; Score 14; DB 13; Length 511;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctgtcttgacag 14
 |||||||||||||
 Db 410 CTTGCTTGAACAG 423

RESULT 15

LOCUS A2249093 513 bp DNA GSS 15-JUN-2000
 DEFINITION RPCI-23-39020.TV RPCI-23 Mus musculus genomic clone RPCI-23-39020,
 DNA sequence.

ACCESSION A2249093
 VERSION A2249093.1 GI:8562296

KEYWORDS GSS.

SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 513)
 Zhao,S., Nieman,W., Feldblum,T., Malek,J., Shatsman,S., Akintet,
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.
 Unpublished (1999)

TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Other GSS: RPCI-23-39020.TV
 COMMENT

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoe@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/Bac_end_intro.html
 Plate: 39 row: 0 column: 20
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..513
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-39020"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI MethyIase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 134 a 98 c 141 g 140 t
 ORIGIN

Query Match 100.0%; Score 14; DB 13; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ctgtcttgacag 14
 |||||||||||||
 Db 394 CTTGCTTGAACAG 407

Search completed: December 8, 2001, 10:21:35
 Job time: 31.75 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 8, 2001, 10:53:12 ; Search time 136.23 Seconds
(without alignments)
23.275 Million cell updates/sec

Title: US-09-508-147-10

Perfect score: 14

Sequence: 1 ctgtctgacagc 14

Scoring table: IDENTITY_NTC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*

2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*

3: /cgn2_6/prodata/2/lna/6A_COMB.seq:*

4: /cgn2_6/prodata/2/lna/6B_COMB.seq:*

5: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:*

6: /cgn2_6/prodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	14	100.0	3688	6	5248670-4
C 2	13	92.9	599	4	US-09-385-982-74
C 3	13	92.9	1446	2	US-08-569-150A-2
C 4	13	92.9	1703	4	US-08-378-313-18
C 5	13	92.9	1869	4	US-09-350-268-1
C 6	13	92.9	2868	3	US-08-274-121B-1
C 7	13	92.9	5224	3	US-08-874-166-46
C 8	13	92.9	7587	4	US-08-378-313-22
C 9	13	92.9	9060	4	US-08-378-313-20
C 10	12	86.6	657	4	US-08-998-416-1061
C 11	12	86.6	1949	4	US-08-818-112-152
C 12	12	86.6	1993	4	US-08-818-112-152
C 13	12	86.6	2088	3	US-09-032-365A-64
C 14	12	86.6	2181	2	US-08-244-205-10
C 15	12	86.6	2181	5	PCT-US92-10284-10
C 16	12	86.6	2220	2	US-08-932-376A-3
C 17	12	86.6	2870	2	US-08-468-036-28
C 18	12	86.6	3046	2	US-08-525-507-7
C 19	12	86.6	3898	1	US-08-471-791-28
C 20	12	86.6	3898	5	PCT-US91-01746-28
C 21	12	86.6	4853	2	US-08-793-824-1
C 22	12	86.6	5285	2	US-08-609-049A-29
C 23	12	86.6	5285	4	US-09-170-996-29
C 24	12	86.6	7573	1	US-08-287-959-2
C 25	12	86.6	7573	1	US-07-846-181-6
C 26	12	86.6	8533	1	US-07-845-989-6
C 27	12	86.6	8533	1	US-07-845-989-6

ALIGNMENTS

C 28	12.4	88.6	12687	1	US-08-676-169-1	Sequence 1, Appl
C 29	12.4	88.6	12687	3	US-08-981-459-1	Sequence 1, Appl
C 30	12.4	88.6	19307	3	US-08-836-022A-10	Sequence 10, Appl
C 31	12.4	88.6	19307	4	US-09-427-048A-10	Sequence 10, Appl
C 32	12.4	88.6	46899	1	US-08-471-119A-1	Sequence 1, Appl
C 33	12	85.7	1518	4	US-09-257-581-4	Sequence 4, Appl
C 34	12	85.7	1518	4	US-09-257-581-6	Sequence 4, Appl
C 35	12	85.7	19307	3	US-08-836-022A-10	Sequence 6, Appl
C 36	12	85.7	19307	4	US-09-427-048A-10	Sequence 7, Appl
C 37	11.4	81.4	22	3	US-09-428-58A-4	Sequence 10, Appl
C 38	11.4	81.4	22	4	US-09-306-998-43	Sequence 4, Appl
C 39	11.4	81.4	30	2	US-08-841-166-8	Sequence 43, Appl
C 40	11.4	81.4	51	2	US-08-841-166-1	Sequence 8, Appl
C 41	11.4	81.4	236	4	US-09-221-298-51	Sequence 1, Appl
C 42	11.4	81.4	239	5	PCT-US93-08106-7	Sequence 51, Appl
C 43	11.4	81.4	239	5	PCT-US93-08106-8	Sequence 7, Appl
C 44	11.4	81.4	239	5	PCT-US94-00089-15	Sequence 8, Appl
C 45	11.4	81.4	239	5	PCT-US94-00089-15	Sequence 15, Appl
					PCT-US94-00089-15	Sequence 16, Appl

RESULT 1
5248670-4/c
Patent No. 5248670.
APPLICANT: DRAWER, KENNETH G.; ECKER, DAVID J.; MIRABELLI, CHRISTOPHER K.; CROOKE, STANLEY T.
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR INHIBITING HERPESVIRUS
NUMBER OF SEQUENCES: 15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/485,297
FILING DATE: 26-FEB-1990
SEQ ID NO: 4:
LENGTH: 3688

Query Match 100.0%; Score 14; DB 6; Length 3688;
Best Local Similarity 100.0%; Pred. No. 13;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgtctgacagc 14
Db 3248 CTGTCTGACAGC 3235

RESULT 2
US-09-385-982-74/c
Sequence 74, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 74
LENGTH: 599
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1)...(599)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-74

Query Match 92.9%; Score 13; DB 4; Length 599;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgcttgacag 14
|||||
Db 103 TTGCTTGACAG 91

RESULT 3
US-08-569-150A-2
Sequence 2, Application US/08569150A
Patent No. 5925804
GENERAL INFORMATION:

APPLICANT: Hoekema, Andreas
APPLICANT: Pen, Jan
APPLICANT: Does, Mirjam P
APPLICANT: Van Den Elzen, Petrus J. M
TITLE OF INVENTION: PRODUCTION OF TREHALOSE IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ladas & Parry
STREET: 26 West 61st Street
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10023

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/4" disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WordPerfect for Windows
SOFTWARE: WordPerfect 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,150A
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP/94/02167
FILING DATE: 30-JUNE-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mass, Clifford J.
REGISTRATION NUMBER: 30,086
REFERENCE/DOCKET NUMBER: U-010552-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 246-1890
TELEFAX: (212) 246-8959
TELEX: No. 5925804e
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1446 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
IMMEDIATE SOURCE:
CLONE: 7f11
POSITION IN GENOME:
MAP POSITION: 41-42'
FEATURE:
NAME/KEY: CDS
LOCATION: 19..1446
OTHER INFORMATION: /product="trehalose phosphate synthase"
OTHER INFORMATION: /gene="otsA"
US-08-569-150A-2

Query Match 92.9%; Score 13; DB 2; Length 1446;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcttgacag 13
|||||
Db 540 CTGCTTGACAG 552

RESULT 4
US-08-378-313-18
Sequence 18, Application US/08378313
Patent No. 620781
GENERAL INFORMATION:

APPLICANT: THEOLOGIS, ATHANASIOS
APPLICANT: SATO, TAKAHIDO
TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/378,313
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/862,493
FILING DATE: 02-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: MORASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 29190-20002.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 856-5600
TELEFAX: (415) 494-0792
TELEX: 706141

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1703 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 11..1489
US-08-378-313-18

Query Match 92.9%; Score 13; DB 4; Length 1703;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcttgacag 13
|||||
Db 1080 CTGCTTGACAG 1092

RESULT 5
US-09-350-268-1/C
Sequence 1, Application US/09350268
Patent No. 6211433
GENERAL INFORMATION:
APPLICANT: Simmons, Carl R.

TITLE OF INVENTION: Methods and Compositions to Enhance Disease Resistance
FILE REFERENCE: In Plants
CURRENT APPLICATION NUMBER: 5718-42A-- M103
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1869
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: M103
US-09-350-268-1

Query Match
Best Local Similarity 92.9%; Score 13; DB 4; Length 1869;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 cttgcttgacag 13
482 CTTGCTTGACAG 470

RESULT 6

US-08-274-121B-1
Sequence 1, Application US/08274121B
Patent No. 6133034
GENERAL INFORMATION:
APPLICANT: Arne Reider Strom
APPLICANT: Ingaf Kaasen
APPLICANT: Olaf Bay Styrvold
APPLICANT: John McDougall
TITLE OF INVENTION: Methods and Compositions
TITLE OF INVENTION: Related to The Production
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1(a)
CURRENT APPLICATION DATA:
FILING DATE: 12-JULY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/893,099
FILING DATE: 27-MAY-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Lassen
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: 34,719
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2868 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: genomic DNA
US-08-274-121B-1

Query Match
Best Local Similarity 92.9%; Score 13; DB 3; Length 2868;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 cttgcttgacag 13
1965 CTTGCTTGACAG 1977

RESULT 7

US-08-874-186-46
Sequence 46, Application US/08874186
Patent No. 5989885
GENERAL INFORMATION:
APPLICANT: Teng, David H-F.
APPLICANT: Tavligian, Sean V.
APPLICANT: Perry III, William L.
APPLICANT: Skolnick, Mark H.
TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE
TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874,186
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,482
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
REFERENCE/DOCKET NUMBER: 24884-121392-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4848
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 5224 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: Intron
LOCATION: 1..1072
FEATURE:
NAME/KEY: exon
LOCATION: 1073..3553
US-08-874-186-46

Query Match
Best Local Similarity 92.9%; Score 13; DB 2; Length 5224;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sat Dec 8 12:41:30 2001

us-09-508-147-10.rni

OY 2 ttgctgaacag 14
 |||||
 Db 3867 TTGCTGACAG 3879

RESULT 8
 US-08-378-313-22
 ; Sequence 22, Application US/08378313
 ; Patent No. 6207881
 ; GENERAL INFORMATION:
 APPLICANT: THEOLOGIS, ATHANASIOS
 APPLICANT: SATO, TAKAHIDO
 TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
 GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: MORRISON & FORSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: California
 ZIP: 94304-1018
 COUNTRY: USA
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/378,313
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/862,493
 FILING DATE: 02-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 29190-20002.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 856-5600
 TELEFAX: (415) 494-0792
 TELEX: 706141
 INFORMATION FOR SEQ. ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7587 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(2637..2813, 2901..3032, 3120..3281, 4540
 LOCATION: ..5106, 5193..5636)
 US-08-378-313-22

Query Match 92.9%; Score 13; DB 4; Length 7587;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgctgaacag 13
 |||||
 Db 5224 CTGCTGACAG 5236

RESULT 9
 US-08-378-313-20
 ; Sequence 20, Application US/08378313
 ; Patent No. 6207881
 ; GENERAL INFORMATION:
 APPLICANT: THEOLOGIS, ATHANASIOS
 APPLICANT: SATO, TAKAHIDO
 TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH

TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: MORRISON & FORSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: California
 ZIP: 94304-1018
 COUNTRY: USA
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/378,313
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/862,493
 FILING DATE: 02-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 29190-20002.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 856-5600
 TELEFAX: (415) 494-0792
 TELEX: 706141
 INFORMATION FOR SEQ. ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810
 LOCATION: ..4376, 4463..4903)
 US-08-378-313-20

Query Match 92.9%; Score 13; DB 4; Length 9060;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgctgaacag 13
 |||||
 Db 4494 CTGCTGACAG 4506

RESULT 10
 US-08-998-416-1061
 ; Sequence 1061, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 APPLICANT: Philippson, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Redischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
 AND USES THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwalis Road
 CITY: Research Triangle Park
 STATE: NO. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO: 1061:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG16460P
US-08-998-416-1061

Query Match 88.6%; Score 12.4; DB 4; Length 657;
Best Local Similarity 92.9%; Pred. No. 93;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
DB 644 CTTGCTTGACAGG 657

RESULT 11
US-08-760-335A-1/c
Sequence 1, Application US/08760335A
Patent No. 5747310
GENERAL INFORMATION:
APPLICANT: SASAKI, Takashi
APPLICANT: SASAKI, Yasuko
APPLICANT: ITO, Yoshiyuki
APPLICANT: OTSU, Kumi
TITLE OF INVENTION: Gene integration into Chromosomes of
Lactobacillus delbrueckii Species and Integrants Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bacon & Thomas
STREET: 625 Slaters Lane - 4th Floor
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,335A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/193,055
FILING DATE: 04-MAR-1994
APPLICATION NUMBER: WO PCT/JP93/055
FILING DATE: 08-JUL-1993

ATTORNEY/AGENT INFORMATION:
NAME: Debenedictis, Joseph
REGISTRATION NUMBER: 28,502
REFERENCE/DOCKET NUMBER: JDB/Sasak1/055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-683-0500
TELEFAX: 703-683-1080
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1949 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactobacillus delbrueckii
US-08-760-335A-1

Query Match 88.6%; Score 12.4; DB 1; Length 1949;
Best Local Similarity 92.9%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cttgcttgacagg 14
|||||
DB 958 CTTGCTTGACAGG 945

RESULT 12
US-08-818-112-152/c
Sequence 152, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ. ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 1993 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: CDS
LOCATION: 152..1273
US-08-818-112-152

Query Match 88.6%; Score 12.4; DB 4; Length 1993;
Best Local Similarity 92.9%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cttgcttgacag 14
|||||
DB 758 CTTGCTTGACAG 745

RESULT 13
US-09-032-365A-64/C
Sequence 64, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502ch, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Nagasari, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
TITLE OF INVENTION: NEUROSENSOR DEFECTS
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032.365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2C1P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 2088 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-032-365A-64

Query Match 88.6%; Score 12.4; DB 3; Length 2088;
Best Local Similarity 92.9%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cttgcttgacag 14
|||||
DB 185 CTTGCTTGACAG 172

RESULT 14
US-08-244-205-10/C
Sequence 10, Application US/08244205
Patent No. 5952544
GENERAL INFORMATION:
APPLICANT: Browse, John, Kinney, Anthony J.,
APPLICANT: Pierce, John, Wierzbicki, Anna M.,
APPLICANT: Yadav, Narendra S., Perez-Grau, Luis
TITLE OF INVENTION: Fatty Acid Desaturase Genes
TITLE OF INVENTION: from Plants
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19898

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244.205
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/804,259
FILING DATE: 4 DECEMBER 1991
ATTORNEY/AGENT INFORMATION:
NAME: Floyd, Linda A.
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1036-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Glycine max
IMMEDIATE SOURCE:
CLONE: pXf1
FEATURE:
NAME/KEY: CDS
LOCATION: 855..1997
US-08-244-205-10

Query Match 88.6%; Score 12.4; DB 2; Length 2181;
Best Local Similarity 92.9%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cttgcttgacag 14
|||||
DB 98 CTTGCTTGACAG 85

RESULT 15
PCT-US92-10284-10/C
Sequence 10, Application PC/TUS9210284
GENERAL INFORMATION:
APPLICANT: Browse, John, Kinney, Anthony J.,
APPLICANT: Pierce, John, Wierzbicki, Anna M.,

```

APPLICANT: Yadav, Narendra S., Perez-Grau, Luis
TITLE OF INVENTION: Fatty Acid Desaturase Genes
TITLE OF INVENTION: from Plants
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10284
FILING DATE: 19921203
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/804,259
FILING DATE: 4 DECEMBER 1991
ATTORNEY/AGENT INFORMATION:
NAME: Floyd, Linda A.
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1036-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4829
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2181 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Glycine max
IMMEDIATE SOURCE:
CLONE: pXf1
FEATURE:
NAME/KEY: CDS
LOCATION: 855..1997
PCT-US92-10284-10

```

```

Query Match      88.6%; Score 12.4; DB 5; Length 2181;
Best Local Similarity 92.9%; Pred. No. 1.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ctgcttgacagg 14
   |||||
DB 98 CTTGCTTGACAGG 85

```

Search completed: December 8, 2001, 10:53:14
 Job time: 4984 sec

THIS PAGE BLANK (USPTO)